UNIVERSITÀ DEGLI STUDI DI MILANO

DOCTORATE SCHOOL IN CHEMICAL SCIENCES AND TECHNOLOGIES

PHD IN CHEMICAL SCIENCE. XXVII CYCLE

Modeling of Molecular Interactions: from Protein-Protein to Halogen Bond

Ph.D. Candidate:
Riccardo De Gonda
Mat: R09692

Supervisor:
prof. Maurizio Sironi

A.A. 2013-2014

Contents

Abstract

Part I - Microtubules are polymeric structures formed by the self association of tubulin dimers. They are extremely dinamical structures, that can undergo phases of growing and shrinking, playing a key role during cells proliferation process. Due to its importance for mitosis, tubulin is the target of many anticancer drugs currently in use or under clinical trial. The success of these molecules, however, is limited by the onset of resistant tumor cells during the treatment, so new resistance-proof compounds need to be developed. We analyze the protein-protein interactions between protofilaments, also known lateral as interaction, using free energy calculations. We were able to identify the most important amino acids for tubulin-tubulin binding and to compare this amino acids with experimental results. The main goal of this study is to perform an atomistic description of the interactions.

Part II - Halogen bond is an important non-covalent interaction which is receiving a growing attention in the study of protein-ligand complexes. Many drugs are halogenated molecules and it has been recently shown that many halogenated ligand establish halogen bonds with biomolecules. Halogen bond is established between an halogn atom and a nucleophilic group due to the presence of a region of positive electrostatic potential, σ-hole. This nucleophilic group can be an atom with lone pairs, for example N,O or S or it can be a system of π electrons of an aromatic ring. This kind of interaction is identified as C-X/π. We developed and implented a method to properly evaluate the halogen bond interaction during a docking simulation using Autodock software. We also study study the effect of substituents on Ph-X/π systems with DFT calculations with two different functional. We identified the best substituent for both rings and compare the results with

Part I

Protein-Protein Interactions

Chapter 1

Computational Methods

1.1 Free Energy

The evaluation of free energies from molecular simulations of biomolecules is extremely important for the purpose of relating microscopic data obtained "in silico" to measurable macroscopic quantities and to understand the physical and structural basis of biological phenomena. Ligand binding to receptors, protein-protein association, protein folding equilibria, transition between different conformations of DNA and transport of small molecules through channels are all processes whose understanding revolves around accurate free energy evaluations. However, despite being one of the most central quantities for biological processes, free energy is also one of the most difficult to compute from computer simulations. Indeed, free energy calculations require a widespread exploration of the configuration space that is extremely expensive, from a computational point of view, resulting generally unfeasible for biomolecules, whose underlying energy landscapes are rugged and complex. Many transformations of biological interest involve conformational changes that span time and/or length scales currently beyond the reach of typical molecular dynamics simulations. To overcome this limitation, many different methods have been developed to calculate free energy changes, whose effectiveness is generally limited to a specific problem. Free energy, entropy and

related quantities are not simple averages of functions of the phase space coordinates. In other words, free energy directly depends on the system partition function and cannot be evaluated from a canonical average.

$$\mu^0_{i,sol} = -RTln\left(\frac{8\pi^2}{C^0}\prod_{i=1}^{M_i}(2\pi m_i RT)^{\frac{3}{2}}\frac{Z_{N,i}}{Z_{N,0}}\right) + P^0\overline{V_i} \qquad (1.1)$$

where

$$Z_{N,i} = \int e^{-\beta U(r_i,r_s)}dr_i dr_s \qquad Z_{N,0} = \int e^{-\beta U(r_s)}dr_s$$

$Z_{N,i}$ $Z_{N,0}$ are called *configuration integrals* and explicitly contain the potential energy $U(r_i, r_s)$ that depends on the solute and solvent coordinates r_i and r_s. These integrals constitute the main difficulty in the calculation of free energies from a simulation, since their evaluation is connected to an efficient sampling of all the conformations involved in the process that, as anticipated, is beyond the reach of molecular dynamics[81]. Indeed, these terms, and free energy in general, depend not only on the lowest energy states, as other mechanical quantities, but on the volume of the conformational space itself.

1.2 Binding Free Energy

Considering the reaction

$$A + B \iff AB$$

where A and B can be either a receptor and a ligand or two proteins involved in the formation of a complex AB. At equilibrium

$$\mu_{AB,sol} = \mu_{A,sol} + \mu_{B,sol}$$

Introducing the standard chemical potential, the general expression for the standard free energy of binding is

$$\Delta G^0_{AB,sol} = \mu_{AB,sol} - \mu_{A,sol} + \mu_{B,sol} = -RTln\frac{\gamma_{AB}C_{AB}C^0}{\gamma_A C_A \gamma_B C_B} = -RTlnK_{AB}$$

Using 1.1, which cancel the kinetic contribution of each species

$$\Delta G^0_{AB,sol} = -RTln\left(\frac{C^0}{8\pi^2}\right)\left(\frac{Z_{N,AB}Z_{N,0}}{Z_{N,A}Z_{N,B}}\right) + P^0\left\langle \Delta V_{AB}\right\rangle \qquad (1.2)$$

where the last term is the pressure-volume work associated with the change in the system size due to the replacement of two free molecule (A,B) by one bound species (AB) and is generally considered to be negligibly small in water at 1 atm. This last term will be accordingly neglected in the following. Since the main term of ΔG^0_{AB} is the ratio of the configuration integrals, direct calculation of the binding energy from molecular dynamics simulation is unfeasible as well. A further simplication can be obtained using implicit solvation the final binding energy can be written as

$$\Delta G^0_{AB} = -RTln\left(\frac{C^0}{8\pi^2}\right)\left(\frac{Z_{AB}}{Z_A Z_B}\right) \qquad (1.3)$$

Many computational techniques have been developed to accurately estimate free energy, binding free energy in particular. The most rigouros of them are based on so-called "alchemical" or structural transformation, such as ***Free Energy Perturbation*** (FEP)[10] or ***Thermodynamics Integration*** (TI)[80]. Since exhaustive phase space sampling makes it possible to compute relative free energies directly, these techniques provide an alternative way to overcome the problem of free energy calculations.

End-points methods

So-called ***End-points*** methods evaluate only the initial and final state of the system. Since they are less computationally expensive, they provide an alternative to the more expensive techniques and are suitable for a greater variety of applications. These approaches are generally

based on a decomposition of the total free energy into different contributions, both enthalpic and entropic. Typically these approaches manipulate 1.3 into the form

$$\Delta G^0_{AB,sol} = \langle U_{AB} \rangle - \langle U_A \rangle - \langle U_B \rangle + \langle W_{AB} \rangle - \langle W_A \rangle - \langle W_B \rangle - T\Delta S^0$$

where U_X and W_X are the (Boltzmann-averaged) potential and solvation energy of species X. More concisely

$$\Delta G^0_{AB,sol} = \langle \Delta U + \Delta W \rangle - T\Delta S^0 \tag{1.4}$$

Molecular Mechanics Generalized Born Surface Area (MM-GBSA), that takes its name from the different approaches used to calculate the different contributions, is among the most used methods of this kind.

1.2.1 MM-GBSA Approach

The MM-GBSA approach[51] uses a thermodynamics cycle (Fig 1.1) to claculate the bindign free energy on the basis of 1.4

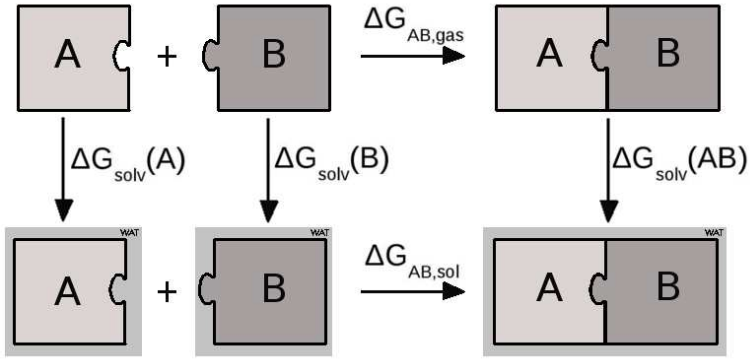


Figure 1.1: Thermodynamic cycle

using the proposed working definition of W(X). Indeed the solvattion contribution can be separated considering the reaction

$$X_{gas} \xrightarrow{\Delta G_{solv}(X)} X_{sol}$$

where $\Delta G_{solv}(X)$ is the free energy of solvation of X

$$\Delta G_{solv}(X) = G(X)_{sol} - G(X)_{gas}$$

the free energy of X in the solvent is then

$$G(X)_{sol} = G(X)_{gas} + \Delta G_{solv}(X)$$

therefore the bindign free energy is.

$$\Delta G(AB)_{sol} = G(AB)_{sol} - G(A)_{sol} - G(B)_{sol}$$

and it is written as

$$\Delta G(AB)_{sol} = \Delta G(AB)_{gas} + \underbrace{\Delta G_{sol}(AB) - \Delta G_{sol}(A) - \Delta G_{sol}(B)}_{\Delta\Delta G_{solv}}$$

decomposing $\Delta G_{AB,sol}$ into a vacuum term $\Delta G_{AB,gas}$ and a solvation term, according to the cycle Fig 1.1. The separation of the two terms is fundamental and allows to calculate the solvation term using implicit solvent models. Theoretically, $\Delta G_{AB,sol}$ can be calculated without the thermodynamic cycle and using an explicit solvent, however the most prominent contributions would result from interactions between the many solvent molecule, with consequent large fluctuations on the final ΔG value (likely to be larger than ΔG itself). Considering that

$$G(X)_{sol} = U(X)_{gas} - TS(X)_{gas} + \Delta G_{solv}(X) \tag{1.5}$$

it can be easily observed that, as compared with 1.4, $\Delta G_{AB,gas}$ loosely corresponds to $\langle \Delta U \rangle - T\Delta S$ while $\Delta\Delta G_{solv}$ corresponds to $\langle \Delta W \rangle$. To calculate the $G(X)_{sol}$ this equation is generally written as

$$G(X)_{sol} = \langle E_{MM} \rangle_X + \Delta G_{solv}(X)$$

and therefore

$$\Delta G_{AB,sol} = \langle E_{MM} \rangle_{AB} - \langle E_{MM} \rangle_A - \langle E_{MM} \rangle_B + \Delta\Delta G_{solv} = \langle \Delta E_{MM} \rangle + \Delta\Delta G_{solv} \quad (1.6)$$

where $\langle E_{MM} \rangle_X$ is the average Molecular Mechanis (MM) energy calculated during the MD simulation

$$\Delta E_{MM} = \underbrace{\Delta E_{bond} + \Delta E_{angles} + \Delta E_{dihedrals}}_{\Delta E_{intra}} + \underbrace{\Delta E_{ele} + \Delta E_{vdw}}_{\Delta Einter}$$

The Solvation Term

While ΔE_{MM} can be easily calculated from a MD simulation in explicit solvent without any further effort, the solvation term must be evaluated "a posteriori" using implicit solvent calculations. The solvation term $\Delta\Delta G_{solv}$ can be decomposend into an electrostatic term and a non-electrostatic term

$$\Delta\Delta G_{solv} = \Delta\Delta G_{elec} + \Delta\Delta G_{non-elec} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} \quad (1.7)$$

where the non-electrostatic term has been further decomposed into Van der Waals term and a "cavity" term, that corresponds to the energy needed to create the cavity occupied by the solute in the solvent.
The first term, the electrostatic one, can be easily computed using the Generalized Born (GB) model[51]. In 1920 Max Born derived the expression for the elecrostatic component of the solvation energy associated with the creation of charge q_i in a spherical cavity with radius a_i in solvent with dielectric constant ϵ

$$\Delta G_{elect} = -\frac{1}{2}\frac{q_i^2}{a_i}\left(1 - \frac{1}{\epsilon}\right)$$

therefore, for N charges q_i, which not overlap cavities, the equation will be

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i}^{N}\frac{q_i^2}{a_i} \tag{1.8}$$

to which an additional term, accounting for the coulombian mutual interaction of charges, when passing from gas phase to the solvent, must be added

$$\underbrace{\frac{1}{2}\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{\epsilon r_{ij}}}_{solv} - \underbrace{\frac{1}{2}\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}}}_{gas} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}}$$

yielding the general equation

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i}^{N}\frac{q_i^2}{a_i} \tag{1.9}$$

However, in a molecule, or in a chemical system in general, we don't have only isolated charges (as ions), but mostly atoms with atomic charges that are connected by bonds. Therefore 1.9 must be modified to take into account this feature of chemical systems

$$\Delta G_{GB} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{f(r_{ij}, a_{ij})}$$

yielding the Generalized-Born equation, where

$$f(r_{ij}, a_{ij}) = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \qquad a_{ij} = \sqrt{a_i a_j} \qquad D = \frac{r_{ij}^2}{(2a_{ij})^2}$$

This equation equals to 1.9 if the charges are well separated since if $r_{ij} >> a_{ij}$

$$D = \frac{r_{ij}^2}{(2a_{ij})^2} >> 1 \rightarrow e^{-D} \approx 0 \rightarrow f(r_{i,a_{ij}}) = r_{ij}$$

alternatively, if $i = j(r_{ij} = 0)$

$$a_{ij} = a_{ii} \rightarrow r_{ij} = 0, D = 0 \rightarrow f(0, a_{ij}) = \sqrt{0 + a_i^2 e^0} = a_i$$

and we obtain 1.8.

The second term, the non-electrostati contribution

$$\Delta G_{nelec} = \Delta G_{cav} + \Delta Gvdw$$

includes a Van der Waals term and a cavity term, ΔG_{cav}, that is substantially the work to create a cavity for the solute in the solvent taking into account the entropy needed to reorganize the solvent molecules around the solute. Since both the reorganization of the solvent and the Van der Waals term are significant only for the first solvation shells, ΔG_{cav} and ΔG_{vdw} can be considered proportional to the surface of the solute exposed to the solvent. For this reason, $\Delta G_{non-elec}$ is generally calculated according to the Solvent Accessible Surface Area (SA) approach and, in particular, as

$$\Delta G_{SA} = \Delta G_{cav} + \Delta G_{vdw} = \gamma A + b \tag{1.10}$$

where A is the totale solvent accessible SA, while γ and B are two constants.

The entropic Term

So far, we have omitted the entropic term in the description of the MM-GBSA approach. However, as explained in section 1.2 the entropy of solvation is included in the calculated solvation term. Hence, of the total entropy

$$\Delta S = \Delta S_{config} + \Delta S_{solv}$$

only the first term is missing. While for some purposes (as Computational Alanine Scanning, see section 1.3) this can be neglected, ΔS_{config} contribution is mandatory to achieve accurate binding energies. Unfortunately, even if the use of end-points approaches allows to separate the free energy in different contributions, avoiding the direct calculation of configurational integrals, ΔS_{config} remains deeply connected to them and difficult to calculate[27;81] . Generally, this term can be written as

$$\Delta S_{config} = \Delta S_{ext} + \Delta S_{int}$$

considering external and internal degrees of freedom, corresponding

$$\Delta S_{config} = \Delta S_{rot+trans} + \Delta S_{vib}$$

where the two terms have been identified as the contribution due to the loss of translational and rotational freedom and to the internal vibration* While $\Delta S_{rot+trans}$ might prevail on ΔS_{vib} in the case of small molecules binding to large receptors, this term is generally neglected for protein-protein association, where the change in vibrational entropy is supposed to be more relevant. ΔS_{vib} can be calculated using normal mode analysis that, however, is too demanding to be applied to an entire protein-protein complex. For all these reason, the inclusion of entropic contributions in free energy calculations is still heterogeneous and, mostly, limited to some approximated terms. In the present work, the inclusion of entropic terms will be limited, especially considering that these terms can be neglected in Computational Alanine Scanning calculations, that are our final purpose.

Appilcation of MM-GBSA

Combining the different contributions, the $\Delta G_{AB,sol}$ is calculated as

*The nature of these contributions and their relative importance is still a matter of debate.

$$\Delta G_{AB,sol} = \langle \Delta E_{MM} \rangle + \Delta G_{GB} + \Delta G_{SA}$$

the main equation behind the MM-GBSA approach. There is no univocal way to proceed in order to apply this equation. The MD simulations can be carried out either in implicit or explicit solvent and, in addition, since we need the different contributions for each species (A, B and the complex AB), distinct simulations of all the species may be performed, or alternatively a simulation of the complex only. While the choice of an implicit solvent might speed up the simulation time, explicit solvation has proved to be a better approximation for biomolecular processes and in particular for protein association. Concerning the simulation instead, even if performing three different MDs for A, B and AB is, theoretically, a better solution, thanks to an advantageous compensation of different errors, the use of a single simulation of the complex yield better results. Calculations for A and B alone are performed "a posteriori" removing one of the two species from the trajectory of the AB complex.

1.3 Computational Alanine scanning

Alanine Scanning (AS) mutagenesis is a valuable procedure to identify important residues ("hot-spots") at the protein-protein interface[23] . It involves the mutation of a generic residue X into an alanine, in order to evaluate the difference in binding energy upon mutation

$$\Delta\Delta G_X = \Delta G_{mut,X} - \Delta G_{wild}$$

where $\Delta G_{mut,X}$ is the free energy of bindigne when X is mutated and ΔG_{wild} is the binding energy of the native protein. Since alanine substitution loosely corresponds to a side-chain removal † it can evaluate the contribution of side-chain functional groups at specific positions ‡.

†the mutated residue retains only the C_β , since alanine has a methyl group attached to the C_α

‡Glycine is theoretically suitable for this purpose being the smallest residue and

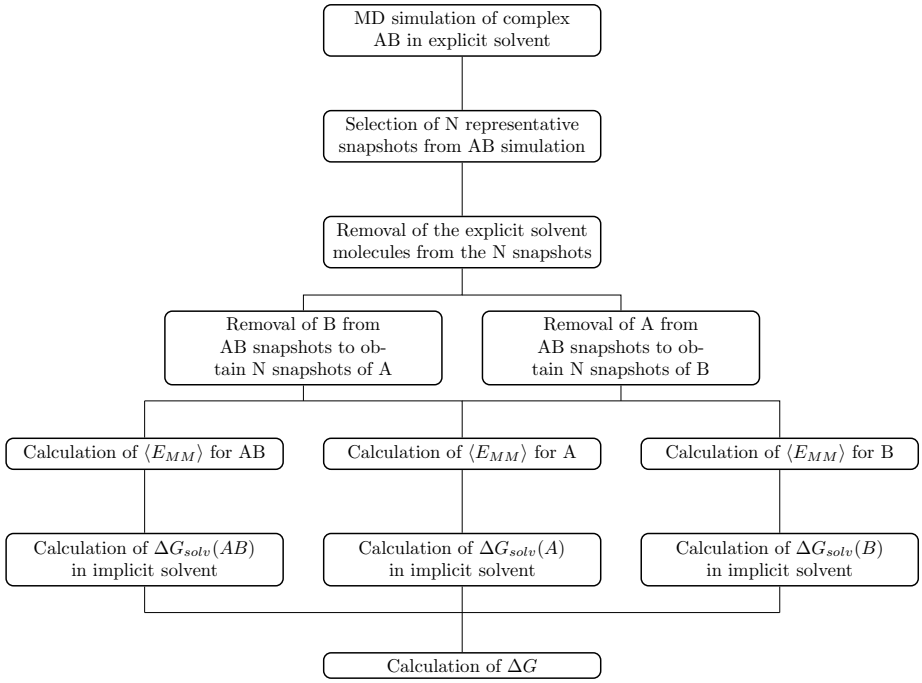


Figure 1.2: MM-GBSA calculation of the free energy of binding

Hot-spots generally cause an increase of the ΔG (a positive $\Delta\Delta G$) of at least 2.0 kcal/mol, corresponding to three order of magnitudes in the binding affinity constant[58]. Experimentally, AS is generally slow and labor-intensive since it requires DNA engineering, protein expression and subsequent purification. Hence, AS is generally applied to map regions that have already been pre-selected; "blind" AS on full surfaces for explorative purposes is generally unfeasible. However, the same technique can be applied "in silico", taking the name of Computational Alanine Scanning (CAS)[50], obtaining reliable results in a much less expensive way, both for what concerns time and resources.
Different protocols exist for CAS and many efforts have been devoted to the development of a fast and accurate method to predict hot-spots[36;24;59]. Generally, the different protocols differ for the simulation conditions, e.g. for the use of one or multiple trajectories for wild-

having no side-chain (just a H atom), but it can introduce conformational flexibility into the protein backbone and therefore is not commonly used.

type and mutated proteins or for the use of an implicit rather than an explicit solvation. Considering the number of MD simulations, three different approaches have been proposed:

single-trajectory CAS Only a simulation of the wild-type protein complex is performed; the structure are then subjected to "post-processing" to obtain the mutated complexes. Monomers are also obtained from the complex MD by removal of the other partner protein from the complex structure, as reported for MM-GBSA calculation.

double-trajectory CAS Two different simulations are performed for the native protein complex and the mutated one; monomers are obtained by "post- processing".

quadruple-trajectory CAS Four MDs are performed for the native complex and the native monomer to be mutated and for the mutated complex and the mutated monomer. The monomer that is not mutated is crossed out by a thermodynamic cycle.

Recently[58] it has been demonstrated that, because of a convenient error cancellation, the use of the **single-trajectory CAS** generally produces better results, in comparison with multiple trajectories approaches, where the error due to insufficient sampling strongly affects the results. Moreover, the use of a single trajectory allows to neglect entropic terms during $\Delta\Delta G$ calculations because of the identity of the structures, leading to mutual deletion of these contributions. Concerning implicit/explicit solvation, there is, unfortunately, little agreement on which approach would achieve better results. While a better description of the solvent is commonly preferred, implicit solvent supporters claim that, since the solvation term ΔG_{solv} in MM-GBSA in calculated using implicit solvent, it is more coherent to use the same method for the MD simulation[§]. Since the largest deviation from experimental results are generally reported for charged residues, Moreira

[§]Both approaches will be used (and compared) in the present work.

et al.[58] have proposed the use of a different internal dielectric constant for different residues in the MM-GBSA calculation[¶]. While better results are obtained using a variable ϵ, we decided to devise a more general solution, consisting in the use of a physiological saline concentration (0.15 M) during free energy calculations. The correction obtained with the latter approach are consistent with those obtained with a variable ϵ.

[¶]Specifically, $\epsilon = 2$ for non-polar residues, $\epsilon = 3$ for polar ones and $\epsilon = 4$ for charged amino acids

Chapter 2

Microtubule

Microtubules (MTs) are important cytoskeletal structures that play a key role in several biological processes, from the formation of cilia and flagella to cell duplication. In particular, MTs form the so-called "mitotic spindle" that separates chromosomes during mitosis, pulling the two sister chromatins to the opposite end of the cell leading to the formation of two daughter cells from a mother one. If MTs are not able to subdivide the genetic material binding chromosomes, the cell is stuck in an intermediate stage of mitosis that, in most cases, degenerates into apoptosis (programmed cells death). Since MTs are crucial for mitosis, they are an ideal target for diseases characterized by an aberrant or unregulated cell duplication, as cancer. While different MT-targeting drugs exist, the problems connected to currently available molecules, including availability and the development of drug- resistance, fuel a constant search for new drugs. Since MTs formation and dynamical behavior is governed by tubulin self-association, the study of tubulin PPIs can a suitable way to develop a new drug.

2.1 Microtubule Structure and Dynamics

Microtubules are hollow cylindrical structures, usually with a 25 nm outer diameter, composed of a lattice of tubulin dimers. Tubulin is a hetero-dimeric protein, composed of an α and a β subunit, with

a high degree of homology. Each subunit is a compact ellipsoid of approximate dimensions 46 x 40 x 65 Å with three domains: a N-terminal one hosting a GTP/GDP binding site, a central small one and a C- terminal prevalently helical domain. Tubulin α, β-dimers stack head-to-tail to form protofilaments that bind laterally to form sheets and are gradually rolled up into a tube-like structure Fig 2.1[41]. MTs can be formed by a variable number of protofilaments from 9 to 16, in physiological conditions is the results of the association of 13 protofilaments[1].

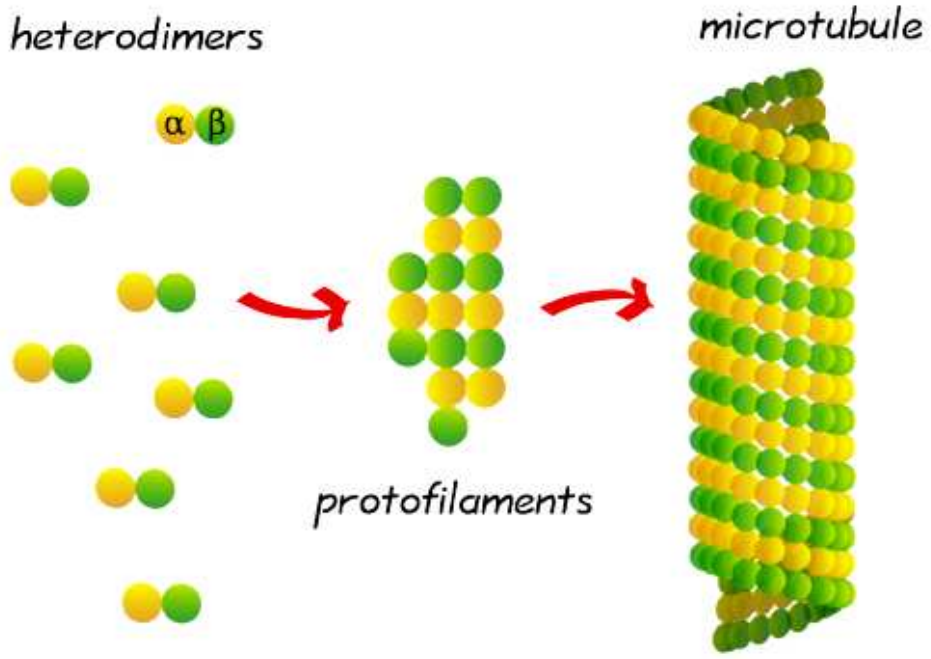


Figure 2.1: Tublulin α, β polymerize to form protofilaments and microtubules.

The final MT structure is organized in a polar manner such that the α-tubulin subunit is exposed at one end, while the β-tubulin subunit is exposed at the other. The two ends of the MT are not equivalent: the β-capped one, called plus-end, is more dynamic, faster growing and faster shrinking than the other, called minus-end[63]. Microtubules polymerization is a complex mechanism, involving nucleation and sub-

sequent elongation, fueled by the energy produced by the hydrolysis of GTP molecules bound to tubulin[83;62]. One molecule of GTP must be associated with both the α and the β subunit for the association of heterodimers to occur. During the addition of a new subunit, the β-terminus of the MT interacts with the α subunit of the new dimer, promoting the hydrolysis of the GTP molecule bound to β-tubulin. The hydrolysis, however, takes place after a certain time from the addition of the dimer, in order to preserve a certain number of GTP-bound subunits at the terminus of the MT, the so-called GTP-cap[77] . After the lag period, the hydrolysis of GTP occurs: the β subunit becomes GDP-bound, while the α subunit continues to bind a molecule of non-hydrolysable GTP Fig 2.2.

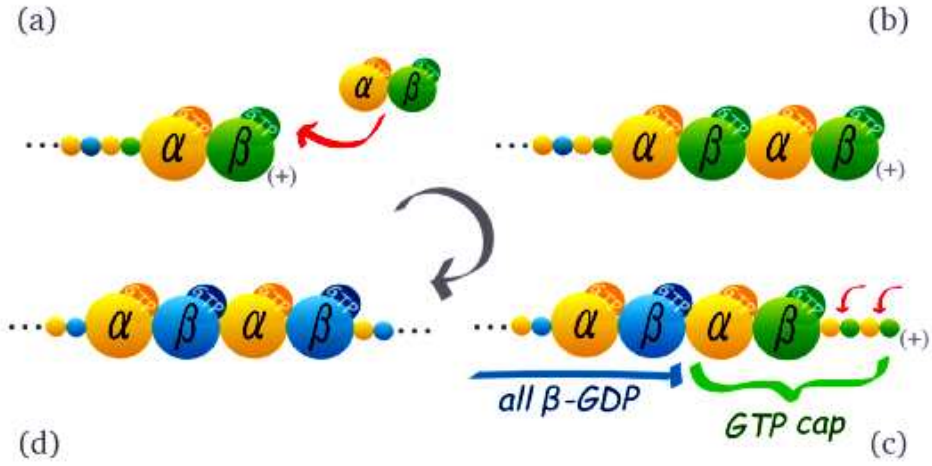


Figure 2.2: Subunits association to form the MT. (a) A new dimer, GTP- bound, can bind to the microtubule plus-end, i.e. the terminal β unit. (b) The "lag" between association and hydrolysis maintains a certain number of GTP-bound units at the plus-end. (c-d) the inner β units progressively hydrolyze the GTP molecule becoming GDP-bound, while α units remain GTP-bound.

2.2 Computational Detail

The system was solvated with TIP3P[42] waters and neutralized with 2964 Na+ ions. The SHAKE algorithm[76] was employed to constraint all bonds involving hydorgem to their equilibrium length, allowing a time step of 1 fs. The system were submittend to 10000 steps of geometry optimization, 1000 steps using the speepest descent and 9000 using the conjugate gradient method. It was then equilibrated for 100 steps with the number of particles, system volume and temperature (300K) constant (NVT conditions) in order to equilibrate temperature of the system, and subsequently for 100 ps with number of particles, system pressure (1atm) and temperature (300K) constant (NPT conditions) in order to equilibrate system density. The equilibration phase was run with 500 kcal mol^{-1} $Å^{-1}$ restraint on the position of the Cα atoms. The system consists of about 7,5 millions of atoms and so perform a molecular dynamics simulation it's too much expensive, so we perform another geometry optimization without restraint on the position of the Cα atoms for 20000 steps to have an equilibrated structure. CAS was carried over all the contact area as described in Cap 1.3.

2.3 Results

2.3.1 Longitudinal Interactions

We perform CAS on all longitudinal surfaces. Table 2.3 shows all hot and warm spots identified on the whole interface. Blues bars refers to $\Delta\Delta G$ value calculated with our protocl, while red bars are those from the previous study[69].

Pieraccini et al. identified 16 hot spots both on α and β subunits. In this work we identified 20 hot spots on the α-subunit and 19 on the β one. On the whole we can say that our protocol is good enough to reproduce the more rigorous protocol reported by Pieraccini et al.. The hot and warm spots identified with our protocol has an energy comparable with the energy of other study in most cases. There is

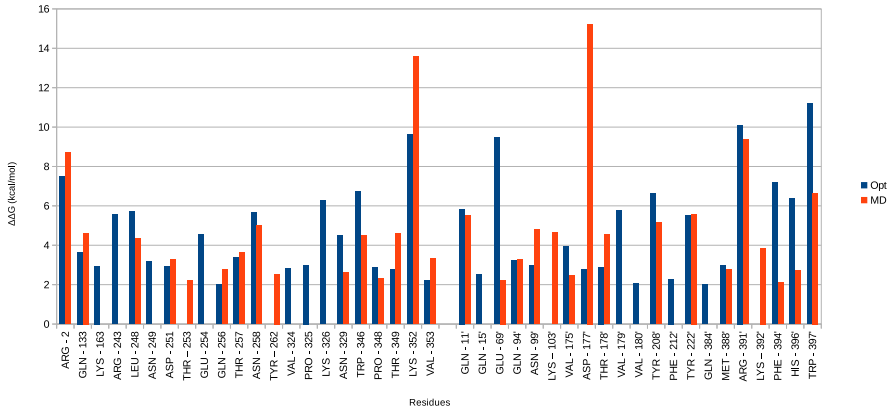


Figure 2.3: Alascan results

some difference, but an in-depth analysis of the causes has not been made because of the time, human and computation, it would require too. For our purpose this results are in good agreement and validate our protocols.

2.3.2 Lateral Interactions

We can identified on the entire contact surface 14 residues, 7 on the right side and 7 on left one that have a key role for lateral interactions. According to the classification described on previous section 1.3 we distinguish residues in hot and warm spots. Table 2.4 report a summary of the whole residues.

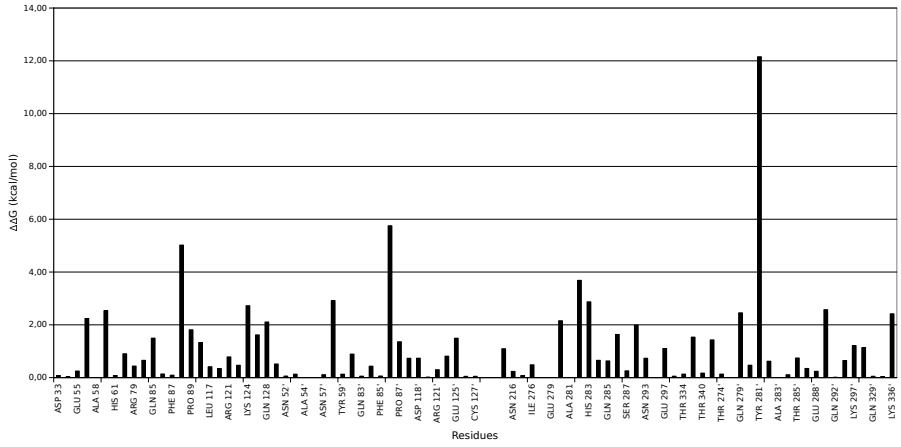


Figure 2.4: Alascan results

We identified 2 hot-spots on left side and 1 on right. The other ones, 5 on right hand and 6 on left, are warm-spots. This residues form a complementary fashion (Figure 2.5) matching electrostatic and hydrophobic features of the two sides.

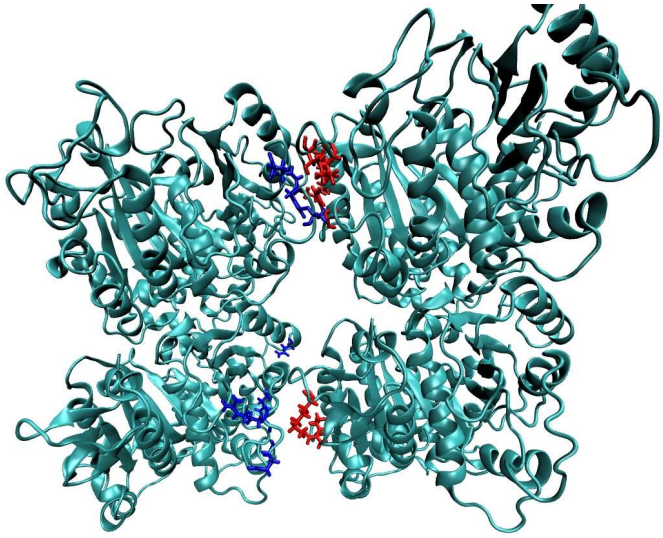


Figure 2.5: Alascan results

A deeper insight into the network of interactions between the residues at the interface that contribute to complex stability can be achieved by

an analysis of the environment of the hot and warm spots. Table 2.1 reports a summary of the residues with $\Delta\Delta G$ and standard errors. The last column specify the side of interaction and the residues marked with the prime belong ti β subunit.

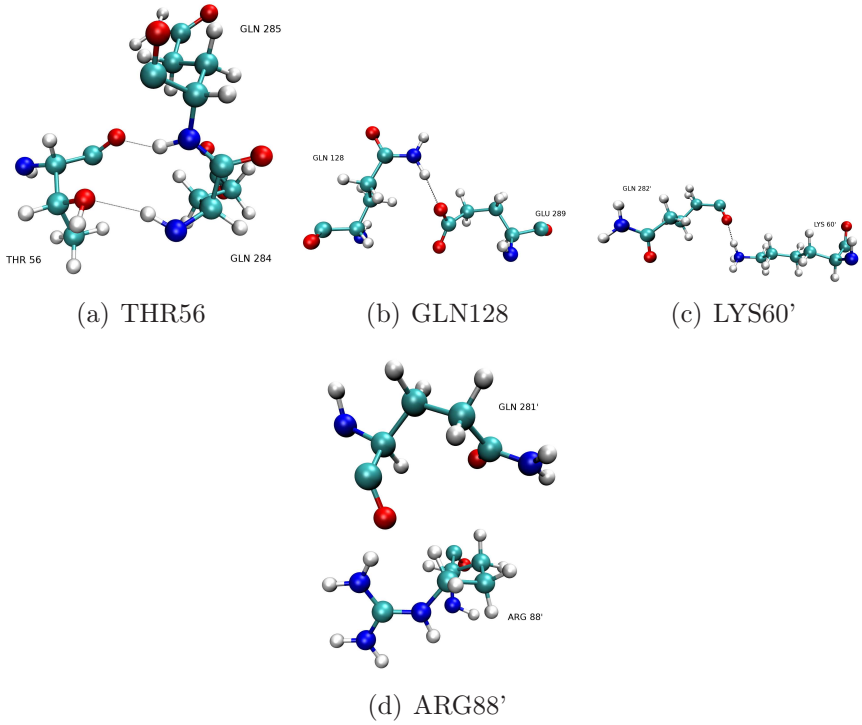
Res Number	Residue	$\Delta\Delta G$	err	side
56	THR	2,25	0,11	L
60	LYS	2,54	0,12	L
88	HIS	5,02	0,17	L
124	LYS	2,73	0,13	L
128	GLN	2,11	0,09	L
58'	LYS	2,92	0,2	L
86'	ARG	5,75	0,19	L
280	LYS	2,16	0,17	R
282	TYR	3,69	0,3	R
283	HIS	2,87	0,17	R
279'	GLN	2,46	0,13	R
281'	TYR	12,16	0,25	R
291'	GLN	2,57	0,13	R
336'	LYS	2,42	0,08	R

Table 2.1: Summary hot and warm spots

Nogales et al.[63] doesn't specify the precise residues involved in interactions, they identify the interactions between loops according to the Lowe et al.[48] classification.

In their work Nogales et al. identify the S7-H9 loop (also known as M-Loop) as the central element of interaction which is in close contact with H3, H2-S3 and H1-S1 loops and they define the character of interaction mainly electrostatic. In our work we were able to make a further step and to go deeply studying the type of interaction and the extact residue involved in the interaction. We identify 3 interaction between M-Loop and H3 loop, 6 between M-Loop and H1-S1 and 3 between M-Loop and H2-S3. We identify also another interaction between H10 loop and H3 loop. Among the whole interactions founded we have 9 electrostatic interactions that correspond to 69% in agreement with

the conclusions of Nogales et al.
Now we show the network of interactions of each residue.
Among the whole ensemble of hot and warm spots we identified 5 hydrogen bond on the left side and one on the right side.



(a) THR56 (b) GLN128 (c) LYS60'

(d) ARG88'

Figure 2.6: H-Bond on the left side

As shown in Fig 2.6(a) THR56 is involved in two hydrogen bonds, but the Alascan can identify only one, with GLU284, because the second involve the backbone oxygen of THR and as already written this part of the aminoacid is not mutated. Due to the two hydrogen bonds we would expected a higher energy, the relatively low energy is due to the fact that the second hydrogen bond is with the backbone, and this kind of interaction is not recognised by Alascan. Fig 2.6(b) and Fig 2.6(c) show the hydrogen bond between GLN128 and GLU289 and between LYS60' and GLN 282'. The energy of this two spots are similar to energy of THR56, in fact also in this cases we have one hydrogen bond detected by Alascan. The last hydrogen bond on this side is Fig 2.6(d).

This is the hydrogen bond between ARG88' and GLN281'. In this case we have a greater energy than other spots, in fact ARG88' is a hot spot, due to the other interaction in which it is inovolved: cation-π interaction.

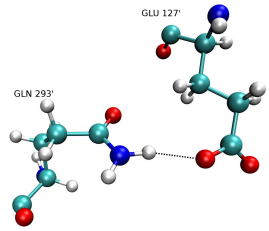


Figure 2.7: H-Bond on the right side

On the other side, the right, we have one hydrogen bond and it is shown in Fig 2.7. This hydrogen bond is between GLN293' and GLU127'. Another important electrostatic interaction is the salt bridge. We found 3 salt bridges.

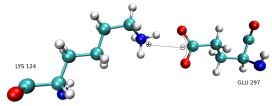
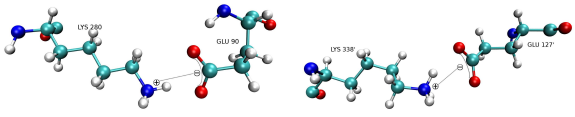


Figure 2.8: Salt Bridge LYS124

We have one salt bridge between LYS124 and GLU 297. The distance between the residues participating in the salt bridge is also cited as being important. The distance required is less than 4 Å. Amino acids greater than this distance do not qualify as forming a salt bridge[44]. In this case the average distance between Nitrogen and the carboxyl group of GLU is 3,05 Å.



(a) LYS280 (b) LYS338'

Figure 2.9: Salt Bridges

We have other salt bridges one between LYS280 and GLU90 and the other between LYS338' and GLU127'. The average distance in the first case is 2,7Å and in the second 2,94Å.
The last two electrostatic interactions identified are one cation-π and one $\pi - \pi$ interactions shown in

(a) ARG88'-TYR283' (b) HIS88-HIS283

Figure 2.10: Salt Bridges on the right side

Fig 2.10(a) shows the cation-π interaction between ARG88' and TYR283', both of these residues are a hot spots. Fig 2.10(b) shows the $\pi - \pi$ interaction between HIS88 and HIS 283, also in this case both are hot spots. We have shown all the electrostatic interactions after a geometric analysis of the hot and warm spots identified by alascan, there is also a network of idrofobic interaction, but thei importance in lateral interaction is limited, as already written and reported in literature[63].

2.4 Conclusion

The problem of this study was the large size of the system. In fact applying the normal protocol (molecular dynamics, CAS) is very difficult and time consuming due to the too huge dimension of the microtubule. Using our protocol we can apply computational alanine scanning to study the lateral interactions. Studying this protein-protein interaface we are able to identify the hot and warm spots. Comparing the results obtained with the experimental ones we identify the same ensemble of residues importat for the lateral interaction and we could perfomr an atomic-level description of the details of the interaction.

Part II

Halogen Bond

Chapter 3

Introduction

3.1 Halogen Bond definition

Halogen bonds Fig 3.1 are particular interactions

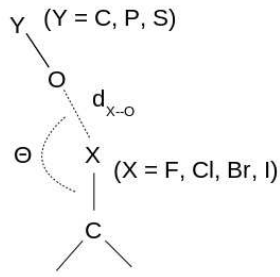


Figure 3.1: Schematic representation of a halogen bond.

established by halogen atoms which present similarities with hydrogen bonds[55;56;47;85;45]. While the hydrogen bond can be schematically represented as $D-H\cdots A$, where D and A denote respectively the donor and the acceptor group (behaving as a Lewis base), the halogen bond can be represented as $D-X\cdots A$ where X is an halogen atom, with the force of the interaction growing up with the dimensions of the halogen (F « Cl $<$ Br $<$ I).

It is well known that hydrogen bonds are of fundamental importance in determining the three-dimensional structure of biomolecules and in molecular recognition processes. For example, α-helices and β-sheets

motifs in proteins, and helices in DNA and RNA are due to a well-defined network of hydrogen bonds. The formation of ligand-protein complexes is also frequently ascribable to the possibility to establish hydrogen bonds between the two partners of the complex. More recently, the importance of halogen bonding in biological systems has been as well evidenced[49;5;64], due in part to the increasing number of resolved protein structures with halogenated ligands. A.R. Voth et al.[4] have also demonstrated that halogen bonds can compete with hydrogen bonds in stabilizing DNA junctions using brominated uracil. It should also be noted that a large number of currently used drugs are halogen-substituted. Halogen atoms are many times introduced in a drug to increase its half-life, so as to prolong its activity in the organism, or to facilitate the membrane permeability so as to permit the drug to reach its biological target. The growing attention towards halogen bonding and a correct interpretation of its nature are now being aimed at increasing the activity of the ligands. For example, a systematic study of the importance of halogen bonding in novel inhibitors of human Cathepsin L has been recently reported[34], showing that it was possible to increase the activity of a ligand of almost two orders of magnitude, by substitution of a hydrogen with an iodine atom. Besides the biological aspects, halogen bonding was also exploited by several research groups as a tool to direct intermolecular recognition processes with applications in crystal engineering[54], in the development of materials with specific electronic properties, ranging from organic semiconductors to superconductors[37;38;43] and also in the development of new materials with non-linear optical properties[18].

3.2 Electrostatic potential anisotropy

A simple interpretation for the formation of halogen bonds was given by Politzer et al.[70;71]. The electrostatic potential of any neutral spherical atom is always positive at any distance from the nucleus, and this can be seen as the predominance of the concentrated nuclear charge

over the dispersed electron cloud. Taking into account the valence electronic structure of a halogen atom covalently bonded to another atom along the z-axis, sketched as $s^2p_x^2p_y^2p_z^1$, it is found that, with respect to the spherical isolated halogen atom (with a 5/3 mean population of p electrons in each direction), there is a depletion of electronic charge along the bond axis, in particular in the region outwards the covalent bond. Such a depletion is compensated by an increase of electronic charge in the directions perpendicular to the bond axis. Based on this interpretation, Politzer was able to explain the strong anisotropy of the electrostatic potential around a covalently bonded halogen atom, as obtained by *ab initio* methods. A region of positive electrostatic potential, called σ-hole Figure 3.2, is in fact built up on the halogen in the region outwards the covalent bond and a crown of negative electrostatic potential is built up around the direction of the D–X bond.

Figure 3.2: σ-hole

Halogen atoms are therefore able to accept electron density from a Lewis base located along the extension of the D–X bond, giving rise to halogen bonding. The size of the positive electrostatic area is found to depend on the electron withdrawing capabilities of the D group and on the nature of halogen atoms. The extension of the σ-hole is in fact more pronounced for the iodine atom, owing to its larger polarizability, decreases for bromine and chlorine atoms, and then disappears for the fluorine atom. Moreover, a positive correlation has been reported between the value of the electrostatic potential on the σ-hole and the strength of the halogen bond[70;71;75].

Chapter 4

Docking

4.1 Introduction

In molecular docking we attempt to predict the structure (or structures) of the intermolecular complex formed between two or more molecules. Docking is widely used to suggest binding modes of protein inhibitors. Nevertheless, using only experimental based techniques, makes it an expensive and time-consuming task. Thats why computational methods were introduced. Docking looks for the correct conformation of a ligand-receptor complex. The most popular programs fit a small molecule to a protein and consider only ligand flexibility with the receptor treated as rigid, because of its size, complexity, and conseguent high-computational cost. Some of this programs have a mixed approach, in fact they treat a small part of the protein, usually some side chains, as flexible. During the last two decades, a large variety of over 60 different docking programs have been proposed for both commercial and academic use. Although they exploit different strategies in the ligand placement, all of them can be categorized into four broad categories:

- **Monte Carlo** also known as Force-Filed methods. They implement Monte Carlo based engine (e.g. Glide[28])
- **Fragment-based** in this approach a ligand is split into frag-

ments which are docked independently and then thei molecule structure is recreated (e.g. Surlfex[39], eHITS[98] and FlexX[73])

- **Evolutionary-based** this apporach use a genetic algorithms to perform the conformational search (e.g GOLD[40] and Autodock[61])
- **Shape complementary** exploits grids to fit the shape of a ligand into an active site of the target combined with Monte Carlo sampling (e.g. LigandFit[84])

In this work we focus our attention on Autdock.

4.2 Autodock Alghoritm

As already written during this work we use only one program for docking: Autodock[61]. This software use a genetic algorithm to generate the ligands conformers in an active site[60]. A Lamarckian genetic alghoritm is employed, where the changes in conformations adopted by molecules after *in situ* optimization are used as a make up for offspring poses. Fig 4.1 shows a schematic flowchart of the entire docking process of the operation.

4.2.1 Scoring function

During the docking procedure, a large number of poses is generated, thus a fast and reliable function that can estimate the free energy of the interaction between the protein and the ligand. It is also crucial to select those conformations that are close to the native structure. Scoring functions express the geometric matching of the two interacting molecules and the strength of this interaction, based on the physicochemical parameters of the system. The main complication of those functions is the estimation of the binding energy as the sum of used terms. Thus, a significant dependence between the size of the ligand and its score can be observed. In fact large molecules, that are able to create many more specific interactions like hydrogen bonds, usually

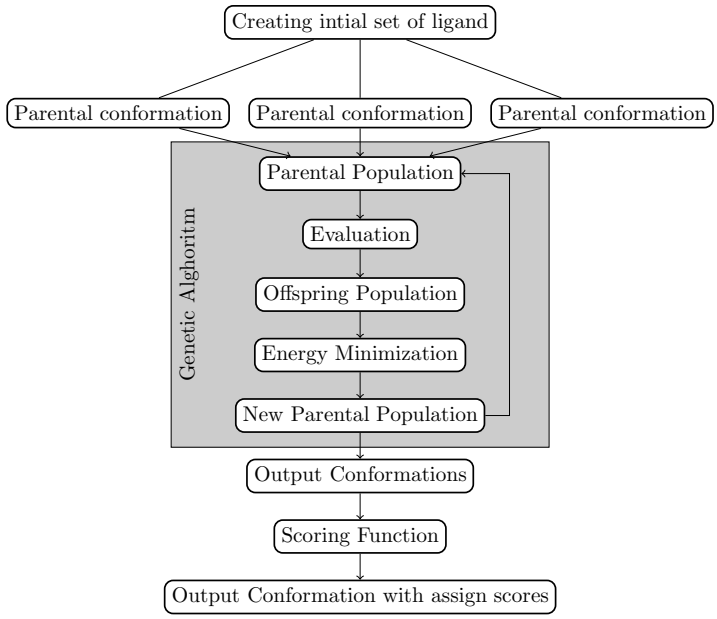


Figure 4.1: Autodock procedure

obtain higher docking score. More than 30 different scoring functions have been developed until 2009[82;97;53;86;68], and they can be grouped in three major categories: force-field based methods[53], solid state empirical methods[26] and knowledge-based (statistical) methods[30].
Autodock uses a semi-empirical free energy force field to evaluate conformations during docking simulations. The force field was parameterized using a large number of protein-inhibitor complexes for which both structure and inhibition constants, or K_i, are known. The semi-empirical free energy force field estimates the energetics of the process of binding of two molecules in a water environment using pair-wise terms to evaluate the interaction between the two molecules and an empirical method to estimate the contribution of the surrounding water. This differs from a traditional molecular mechanics force field, which also relies on pair-wise atomic terms, but typically uses explicit water molecules to evaluate solvation contributions.
The approach taken in AutoDock is shown in Fig 4.2. The free energy of binding is estimated to be equal to the difference between (1) the

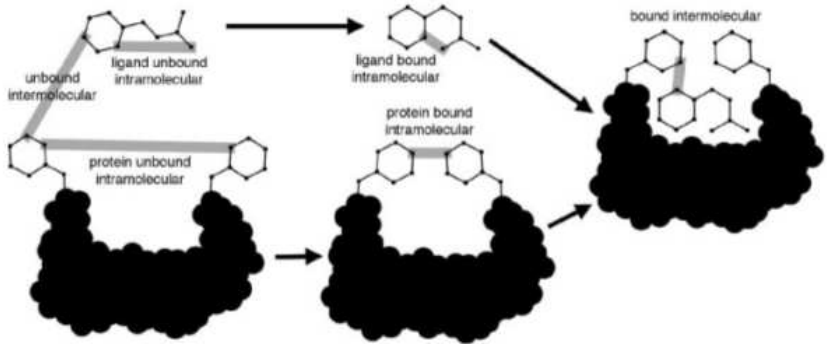


Figure 4.2: Free Energy of Binding evaluation process

energy of the ligand and the protein in a separated unbound state and (2) the energy of the ligand–protein complex. This is broken into two steps: we evaluate the intramolecular energetics of the transition from the unbound state to the bound conformation for each of the molecules separately, and then evaluate the intermolecular energetics of bringing the two molecules together into the bound complex. The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy loss upon binding (ΔS_{conf}):

$$\Delta G = (V^{L-L}_{bound} - V^{L-L}_{unbound}) + (V^{P-P}_{bound} - V^{P-P}_{unbound}) + (V^{P-L}_{bound} - V^{P-L}_{unbound} + \Delta S_{conf}) \tag{4.1}$$

In this equation, L refers to the "ligand" and P refers to the "protein" in a protein–ligand complex; note, however, that the approach is equally valid for any types of molecules in a complex. The first two terms are intramolecular energies for the bound and unbound states of the ligand, and the following two terms are intramolecular energies for the bound and unbound states of the protein. The change in intermolecular energy between the bound and unbound states is in the third parentheses. It is assumed that the two molecules are sufficiently distant from one another in the unbound state that $V^{P-L}_{unbound}$ is zero. In the current study, we did not allow flexibilty in the protein, so the bound state of the protein is identical with the protein unbound state,

and the difference in their intramolecular energy is zero. The pair-wise atomic terms include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{i,j} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (4.2)$$

The weighting constants W are the ones that are optimized to calibrate the empirical free energy based on a set of experimentally characterized complexes. The first term is a typical 6/12 potential for dispersion/repulsion interactions. Parameters A and B were taken from the Amber force field[87].The second term is a directional H-bond term based on a 10/12 potential[31]. The parameter C and D are assigned to give a maximal well depth of 5 kcal/mol at 2.5 Å for O-H and N-H, and a depth of 1 kcal/mol at 2.5 Å for S-H as shown in eq 4.3.

$$C_{ij} = \frac{\epsilon_{ij}}{(12-10)} * R_{ij}^{12} * 10 \qquad D_{ij} = \frac{\epsilon_{ij}}{(12-10)} * R_{ij}^{10} * 10 \quad (4.3)$$

Directionality of the hydrogen bond interaction $E(t)$ is dependent on the deviation of th angle t from ideal bonding geometry and is described fully in works of Morris et al. and Boobbyer et al.. Electrostatic interactions are evaluated with a screened Coulomb potential[52]. The final term is a desolvation potential based on the volume (V) of the atoms surrounding a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance[79]. The distance weighting factor σ is set to 3.5 Å. The term for the loss of torsional entropy upon binding (ΔS_{conf}) is directly proportional to the number of rotatable bonds in the molecule (N_{tors}):

$$\Delta S_{conf} = W_{conf} N_{tors} \quad (4.4)$$

4.3 Computational Methods

In this work, as already written in previous section, we use Autodock 4.2 fore docking calculation. Usually Autodock can't identify halogen bond interaction, see 4.2.1. In order to perform a docking simulation of an halogenated ligand and can reproduce correctly the halogen bond we developed a new strategy. Usually during an Autodock calculation each atom is defined by this parameters:

- Atom type
- R_{ii} = sum of van der Walls (vdW) radii of two like atoms (Å)
- epsii = vdW well depth (kcal/mol)
- atomic solvation volume (Å^3)
- R_{ii-hb} = H-bond distance between heteroatomand hydrogen (Å). Is set zero for hydrogen because the value is included in the heteroatom record.
- $epsii_{hb}$ = well depth for hydrogen bond (kcal/mol)
- hbond = integer indicating the type of hbond:

  0 no hbond

  1 spherical H donor

  2 directional H donor

  3 spherical acceptor

  4 directional N acceptor

  5 directional O/S acceptor

Autodock has its default parameter file, and it automatically assigns to each atom an atom type. Subsequently we perform the docking calulation and for each run we obtain one structure. Usually a good docking calculation perform at least 100 run, and make a cluster analysis of the

100 structures obtained. The structure with lower energy of the most populated cluster is likely to reproduce the native docked conformation. Due to the great similarity with hydrogen bond (3.1) we model our approach to that used to describe hydrogen bond in Autodock. As was done in previuos work[74] we introduce a pseudo-atom with a positive charge to mimick the σ-hole. This pseudo-atom is a zero-volume interaction point. The parameters for this atoms are:

```
atom_par Ps  0.00 0.000 00.0000 -0.00000 0.0 0.0 2 -1 -1 3
```

Ps is an atom with no volume, no van der Walls well depth and no atomic solvation volume. It can be involved in a directionl bond similar to a H-bond. The value of the bond distance and of the well depth of bond is included in the heteroatom (O, N, S).
We test this protocol on a ensemble of structures present in the Protein Data Bank. To identify the structures we use Relibase (v 3.0)[35]. Relibase is a web-based system for searching and analysing protein-ligand structures in the Protein Data Bank (PDB)[9]. Relibase search structures which fit with geometrical parameters chosen by the user. We choose parameters which fit with the directionality of the halogen bond. We choose $\Theta > 130°$ and $d < 3.5$Å*. Using Relibase we identify 14 structures and we have performed docking simulation with and without pseudoatom.

*see Figure 3.1 for symbol references

4.4 Results

4.4.1 1ZOE

The crystallographic structure has one halogen bond:

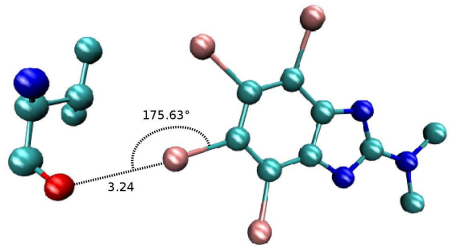
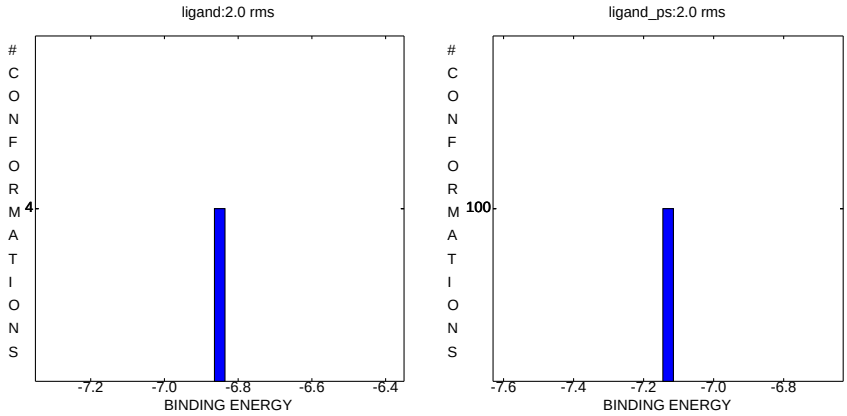


Figure 4.3: Halogen bonds in crystallographic strutucre

The parameters of this one halogen bond are:

residue	d	Θ
VAL116	3.24	175.63

Table 4.1: Crystallographic parameters



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.4: Cluster analysis with and without pseuatom

Figure 4.4 shows the cluster analysis. For both cases we have one cluster. Checking the parameters of the two cases when we use the

structure without pseudoatom Autodock can't reproduce the correct geometry, when we introduce the pseudoatom and check the parameters this are in agreement with the crystallographic ones, as shown in Table 4.2

res	d	Θ	d	Θ
VAL116	—	—	3.18	174.06
	without PS		with PS	

Table 4.2: Parameters after docking simulation

If we taking into accout only this parameters we could say that the introducing of pseudoatom reproduce the halogen bond, but if we check which bromine atom is involved in the halogen bond we see that it is the wrong atom.

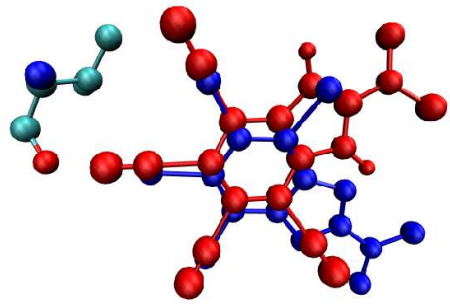


Figure 4.5: Superimposion of crystallographic structure (blue) and structure with pseudoatom (red)

Figure 4.5 shows a superimposition of the conformation found after docking simulation with the pseudoatom and the crystallographic structure. We can see that the structure is slightly rotated and the bromine atom that form the halogen bond is different.

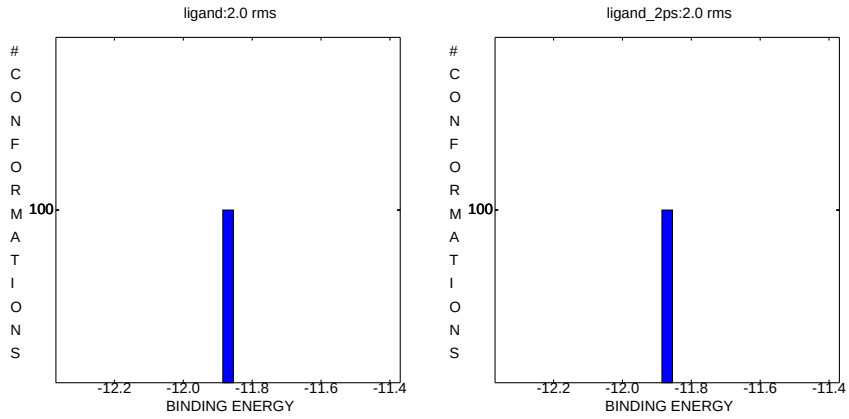
4.4.2 2J4A

This structure has one halogen bond. The paramenters of the halogen bond for this structure is shown in Table 4.3.

res	d	Θ
PHE272	3.28	163.62

Table 4.3: Crystallographic parameters

Figure 4.6 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.6: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
PHE272	3.73	167.88	3.45	167.97
	without PS		with PS	

Table 4.4: Parameters after docking simulation

Table 4.4 shows the parameters for the two simulation. Whithout pseudoatom the distance is too high so only with the pseudoatom the correct conformation.

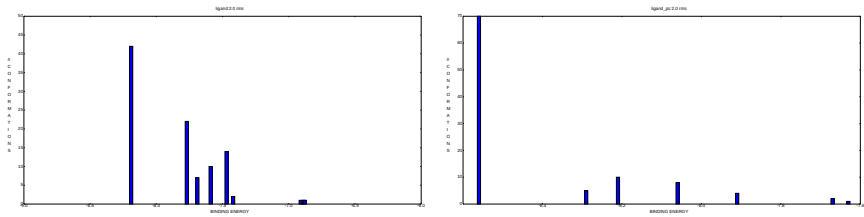
4.4.3 2QLQ

This structure has one halogen bond. the parameters of this halogen bond are shown in Table 4.5.

res	d	Θ
ILE336	3.32	165.75

Table 4.5: Crystallographic parameters

Figure 4.7 shows the cluster analysis of the two docking simulation.



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.7: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
ILE336	5.93	45.24	5.92	45.87
without PS			with PS	

Table 4.6: Parameters after docking simulation

Table 4.6 shows the parameters for the two simulation. In both cases the distace is too high and the angle is too low. So docking simulation doesn't reproduce the halogen bond in both cases.

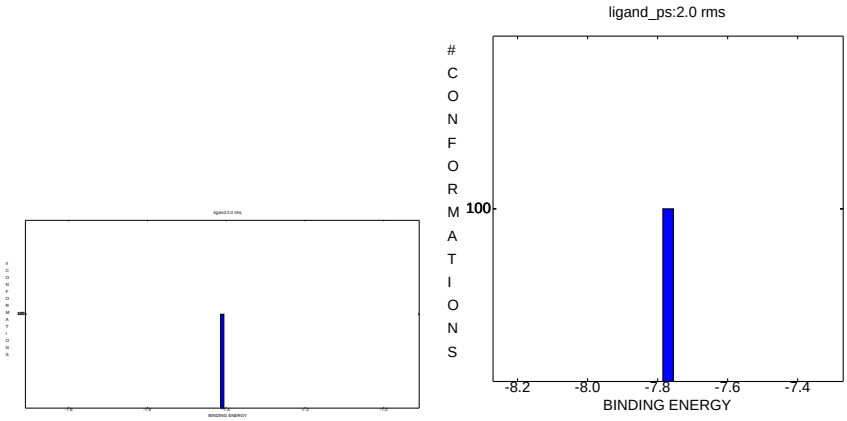
4.4.4 2VQS

The halogen bond of this structure is:

res	d	Θ
SER106	3.39	173.18

Table 4.7: Crystallographic parameters

Figure 4.8 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.8: Cluster analysis with and without pseuatom

d	Θ	d	Θ
3.10	132.51	3.11	166.71
without PS		with PS	

Table 4.8: Parameters after docking simulation

Table 4.8 shows the parameters for the two simulation. Both simulations reproduce the halogen bond, but when we introduce the pseudoatom we have an improvement of the result, in fact the angle became more similar to the crystallographic one.

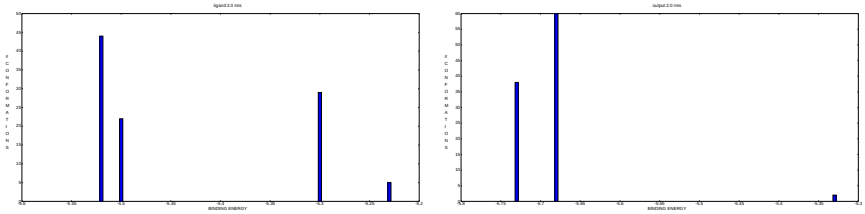
4.4.5 4AJE

The halogen bond of this structure is:

d	Θ
3.31	162.56

Table 4.9: Crystallographic parameters

Figure 4.9 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.9: Cluster analysis with and without pseuatom

d	Θ	d	Θ
—	—	3.26	161.45
without PS		with PS	

Table 4.10: Parameters after docking simulation

Table 4.10 shows the parameters for the two simulation. Simulation with pseudoatom reproduce the halogen bond, without pseudatom instead the geometry is too different tp check the parameters.

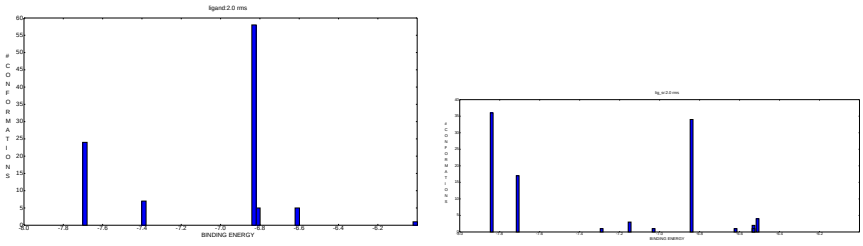
4.4.6 2QS1

The halogen bond of this structure is:

d	Θ
3.41	164.01

Table 4.11: Crystallographic parameters

Figure 4.10 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.10: Cluster analysis with and without pseuatom

d	Θ	d	Θ
—	—	3.46	164.85
without PS		with PS	

Table 4.12: Parameters after docking simulation

Table 4.12 shows the parameters for the two simulation. Simulation with pseudoatom reproduce the halogen bond, without pseudatom instead the geometry is too different tp check the parameters.

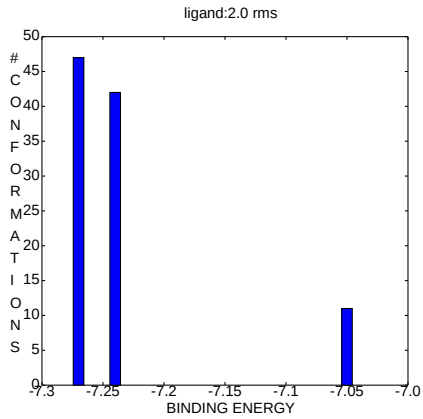
4.4.7 3D1G

The halogen bond of this structure is:

res	d	Θ
thr172	3.02	168.56
val360	3.46	166.26

Table 4.13: Crystallographic parameters

Figure 4.11 shows the cluster analysis of the two docking analysis



(a) With PseudoAtom

Figure 4.11: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
thr172	—	—	—	—
val360	—	—	3.13	164.85
	without PS		with PS	

Table 4.14: Parameters after docking simulation

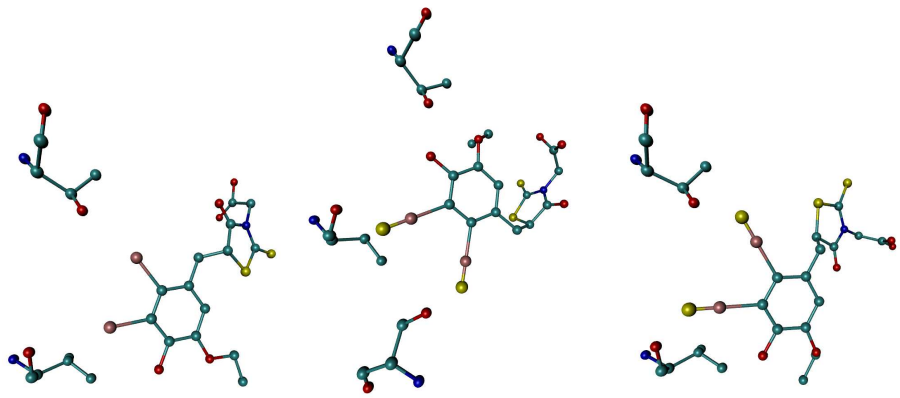
Table 4.14 shows the parameters for the two simulation. This case is interessant. In fact without pseudoatom we have only one cluster and the geometry don't reproduce the halogen atom. When we check the cluster analysis of simulation with the pseudoatom we have three clusters. Usually we would chose the most populated cluster, so we

would have chosen the first one. This conformation reproduce only one halogen bond. In this case the position of the bromine atom involved in the halogen bond is right, but the conformation isn't the right one, in fact this structure is simmetric respect the axis of the main chain of the molecule. If we take a look to the conformation of the second cluster, this structure reproduce both the halogen bonds s shown by the parameters in Table 4.15.

res	d	Θ
thr172	3.14	150.88
val360	3.12	147.55
	with PS	

Table 4.15: Parameters of the structure of the second cluster

The Figure 4.12 show the crystallographic structure (a) and the conformation of the first cluster (b) and the confromation of the second one (c). As we can see the second cluster reproduce better the right geometry, but on the basis of the cluster analysis we would have chosen the first one.



(a) (b) (c)

Figure 4.12: Comparison between geometry

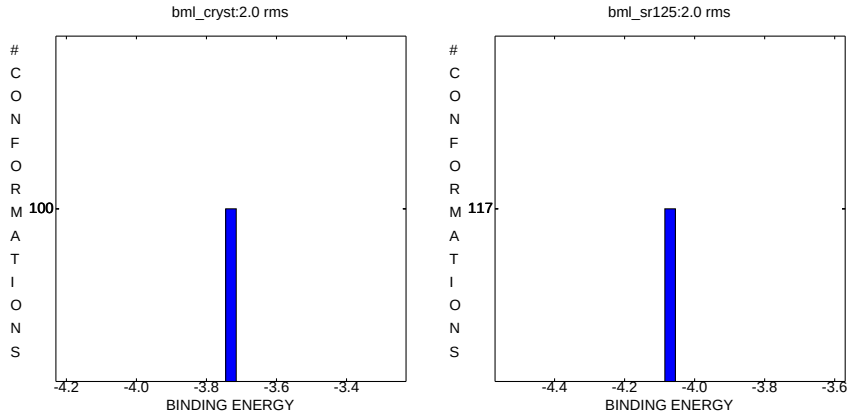
4.4.8 BML

The halogen bond of this structure is:

d	Θ
3.30	171.40

Table 4.16: Crystallographic parameters

Figure 4.10 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.13: Cluster analysis with and without pseuatom

d	Θ	d	Θ
without PS		with PS	

Table 4.17: Parameters after docking simulation

Table 4.17 shows the parameters for the two simulation. Both cases don't reproduce the halogen bond.

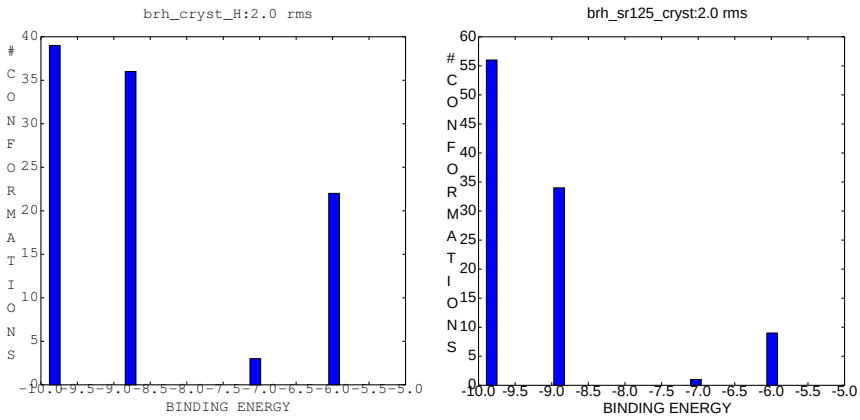
4.4.9 BRH

The halogen bond of this structure is:

res	d	Θ
GLU13	3.26	167.68

Table 4.18: Crystallographic parameters

Figure 4.14 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.14: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
GLU13	4.71	128.87	4.58	130.62
without PS			with PS	

Table 4.19: Parameters after docking simulation

Table 4.17 shows the parameters for the two simulation. Both cases don't reproduce the halogen bond.

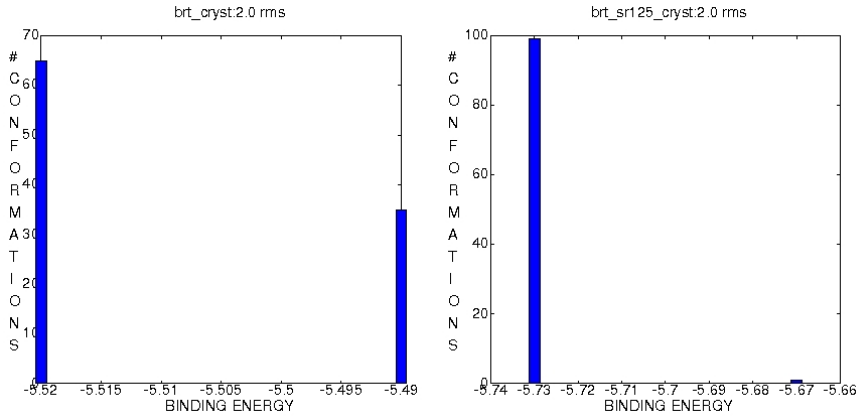
4.4.10 BRT

The halogen bond of this structure is:

d	Θ
3.15	

Table 4.20: Crystallographic parameters

Figure 4.15 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.15: Cluster analysis with and without pseuatom

d	Θ	d	Θ
3.81	—	3.06	—
without PS		with PS	

Table 4.21: Parameters after docking simulation

Table 4.21 shows the parameters for the two simulation. The docking simulation without pseudoatom don't reproduce the halogen bond, the structure with pseudoatom instead can reproduce the alogen bond.

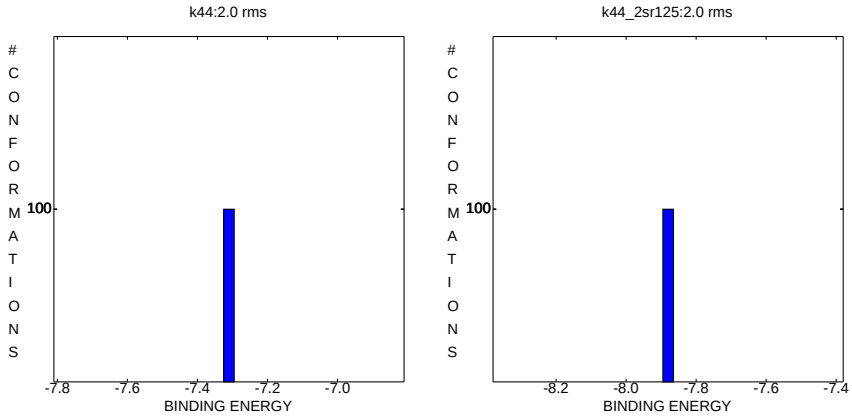
4.4.11 K44

The halogen bond of this structure is:

res	d	Θ
GLU109	3.18	1544.56
VAL111	2.89	177.45

Table 4.22: Crystallographic parameters

Figure 4.15 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.16: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
GLU109	3.44	129.89	3.07	139.39
VAL111	3.13	171.51	3.00	167.74
without PS			with PS	

Table 4.23: Parameters after docking simulation

Table 4.23 shows the parameters for the two simulation. When we perfomr the docking simulation of the structure without the pseudoatom Autodock can reproduce only one halogen bond (Br-VAL111). When we introduce the peudoatom the halogen bond is reproduced.

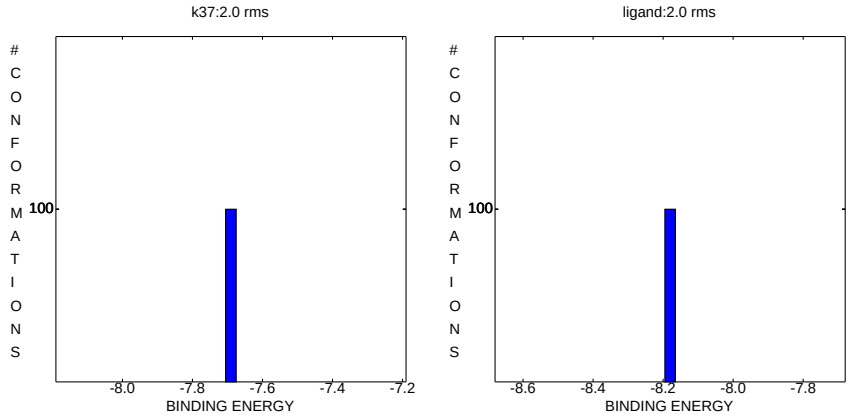
4.4.12 K37

The halogen bond of this structure is:

res	d	Θ
VAL111	2.98	175.35
GLU114	3.23	143.60

Table 4.24: Crystallographic parameters

Figure 4.17 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.17: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
VAL111	—	—	3.09	169.72
GLU114	—	—	3.10	140.31
without PS			with PS	

Table 4.25: Parameters after docking simulation

Table 4.25 shows the parameters for the two simulation. The simulation without pseudoatom don't reproduce the halogen bond in neither case, when we introduce the pseudoatom we reproduce both halogen bonds.

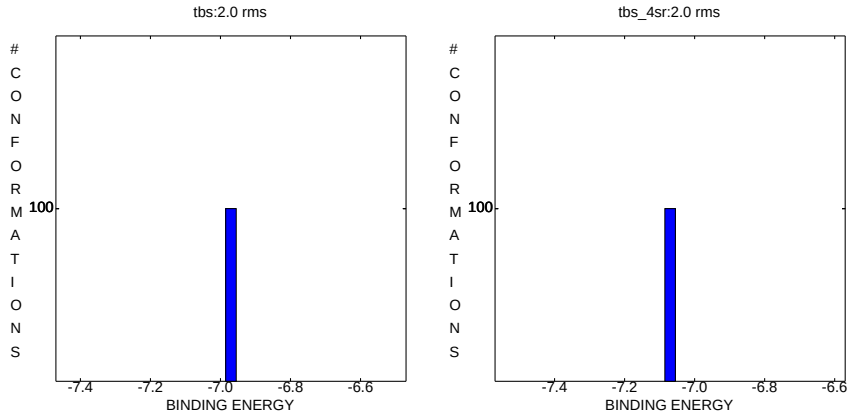
4.4.13 TBB

The halogen bond of this structure is:

res	d	Θ
ILE11	3.24	144.08
LEU84	2.90	164.70
GLU82	3.01	168.55

Table 4.26: Crystallographic parameters

Figure 4.18 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.18: Cluster analysis with and without pseuatom

res	d	Θ	d	Θ
ILE11	—	—	2.95	132.41
LEU84	—	—	3.12	100.34
GLU82	—	—	—	—
without PS			with PS	

Table 4.27: Parameters after docking simulation

Table 4.27 shows the parameters for the two simulation. Both simulations don't reproduce the halogen bonds.

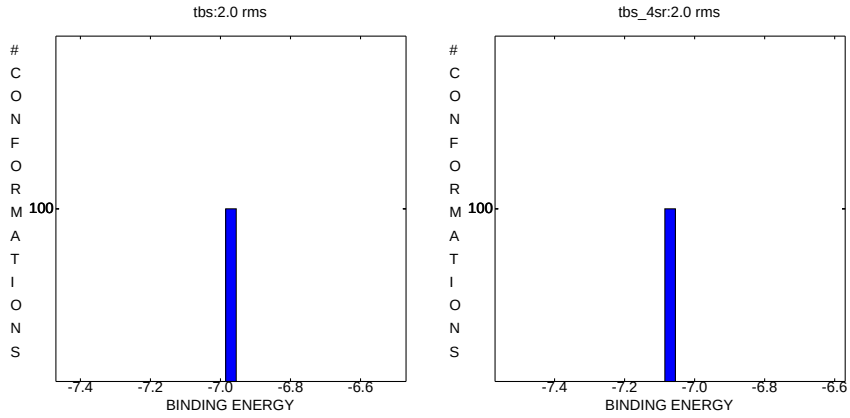
4.4.14 TBS

The halogen bond of this structure is:

d	Θ
2.99	165.33

Table 4.28: Crystallographic parameters

Figure 4.19 shows the cluster analysis of the two docking analysis



(a) Without PseudoAtom (b) With PseudoAtom

Figure 4.19: Cluster analysis with and without pseuatom

d	Θ	d	Θ
—	—	5.61	141.28
without PS		with PS	

Table 4.29: Parameters after docking simulation

Table 4.27 shows the parameters for the two simulation. Both simulations don't reproduce the halogen bonds.

4.4.15 Discussion

PDB	without pseudo	with pseudo
1ZOE	no	no
2J4A	no	yes
2QLQ	no	no
2VQS	yes	yes
4AJE	no	yes
2QS1	no	yes
3D1G	no	no
BML	no	no
BRH	no	no
BRT	no	yes
K44	no	yes
K37	no	no
TBB	no	no
TBS	no	no

Table 4.30: Overview of results.

Table 4.30 shows an overview of the results. "no" means that the docking simulation don't reproduce the halogen bond, "yes" the opposite. On the whole without pseudoatom Autodock reproduce the halogen bond only one time, with the pseudoatom instead 6 times. On the whole we reproduce the halogen bond and the correct crystallographic strucutre 43% times. In conclusion we can say that we make a first step in the implementation in Autodock of a new approach to reproduce halogen bond, but more work is needed to fine tune the parameters of pseudoatom.

Chapter 5

Density Functional Theory

5.1 The mathematical foundation of DFT

The most simple Schrödinger equation is easily written for the nonrelativistic, time-independent, 1-particle case:

$$\left[-\frac{\hbar^2\nabla^2}{2m}+\nu(r)\right]\Psi(r)=\epsilon\Psi(r) \tag{5.1}$$

A convenient starting point for a mathematical derivation of density functional theory of atoms and molecules, is the extension of 5.1 to treat N-particles systems:

$$\left[\sum_i^N\left(-\frac{\hbar^2\nabla_i^2}{2m}+\nu(r_i)\right)+\sum_{i<j}U(r_i,r_j)\right]\Psi(r_1,r_2,...,r_N)= E\Psi(r_1,r_2,...,r_N) \tag{5.2}$$

in compact way it looks like:

$$[\hat{T}+\hat{U}+\hat{V}]\Psi(r_1,r_2,...,r_N)=E\Psi(r_1,r_2,...,r_N) \tag{5.3}$$

where $\hat{T}$ is the kinetic energy operator, $\hat{U}$ is tho Coulomb electron-electron interaction potential and finally $\hat{V}$ is the electron-nuclei interaction potential:

$$\hat{T} = -\frac{\hbar^2}{2m}\sum_i \nabla_i^2 \qquad \hat{u} = \sum_{i<j} U(r_i, r_j) = \sum_{i<j} \frac{q^2}{|r_i - r_j|} \qquad \hat{V} = \sum_i \nu(r_i)$$

It is interesting to note here that the first two operators are universal: $\hat{T}$ is in fact the same for all nonrelativistic systems, and $\hat{U}$ is analogous for any systems of particles interacting through a Coulomb potential. Whether our system is an atom, a molecule, or a solid thus depends only on the term $\nu(r_i)$, that ususally it is called external potential, and sometimes it is denoted also $\hat{V_{ee}}$. For examples for atoms it is:

$$\hat{V} = -\sum_i \frac{Z_e}{|r_i - R|}$$

while for molecules and solids:

$$\hat{V} = -\sum_{ik} \frac{Z_k e^2}{|r_i - R_k|}$$

The usual quantum-mechanical approach to Schrödinger equation is to specify the system by chosing $\nu(r)$, solving the Schrödinger equation for the wavefunction, and then calculates observable through this wavefunction. Among the observables that are calculated in this way, is the particle density:

$$\rho(r) = N \int dr_2 .... \int dr_N \Psi * (r_1, r_2 ... r_N) \Psi(r_1, r_2 ... r_N) \tag{5.4}$$

The density functional approach invert the above procedure: from the electron density $\rho(r)$, one can calculate the wavefunction, and from there, all other observables. This main idea is supported by the fundamental theorems of DFT: the Hohenberg-Kohn theorems, described in the next section.

5.1.1 Hohenberg-Kohn theorems

At the heart of DFT is the first Hohenberg-Kohn (H-K) theorem, that states for ground-state densities that equation 5.4 can be inverted to obtain ground-state wavefunctions. In simple words, given a ground-state density $\rho_0(r)$, it is possible to obtain the corresponding ground-state wavefunction $\Psi_0(r_1, r_2, ...r_N)$, consequently all ground-state observables are simple functionals of the electron density $\rho_0(r)$:

$$O[\rho_0] = \left\langle \Psi[\rho_0] \middle| \hat{O} \middle| \Psi[\rho_0] \right\rangle \tag{5.5}$$

The second H-K theorem proves that the ground-state electron density minimizes the energy functional:

$$E_\nu[\rho] = \min_{\Psi \to \rho} \left\langle \Psi \middle| \hat{T} + \hat{U} + \hat{V} \middle| \Psi \right\rangle \qquad E_\nu[\rho_0] \leq E_\nu[\rho'] \tag{5.6}$$

This theorem provides a variational principle for the electron density, and is crucial for all the further development of DFT. In the same way as in 5.3, $\hat{T}$ and $\hat{U}$ are universal functionals, while $\hat{V}$ is a non-universal functional that depends on the system that can be written explicitly in terms of the particle density ρ:

$$V[\rho] = \int dr \nu(r)\rho(r) \tag{5.7}$$

Although the Hohenberg-Kohn theorems are extremely powerful, they do not offer a practical way of computing the ground-state density of a system. In principle the only thing that must be done is the minimization of the energy functional $E_\nu[\rho]$ with respect to $\rho(r)$:

$$E_\nu[\rho] = T[\rho] + U[\rho] + \int dr \nu(r)\rho(r) \tag{5.8}$$

In practice, the exact functional is not known and we need to used approximations for $T[\rho]$ and $V[\rho]$. The first example of density-functional calculation was the Thomas-Fermi model, but the most modern and useful approach to DFT calculations is the Kohn-Sham formulation described in the next section.

5.1.2 Kohn-Sham DFT

The ground-state energy of a many-electron system can be obtained by minimising the energy functional 5.8, subject to the constraint that the number of electrons N is conserved. Using Lagrange multipliers, 5.8 becomes:

$$\delta\left[T[\rho]+U[\rho]+\int dr\nu(r)\rho(r)-\mu\left(\int dr\rho(r)-N\right)\right]=0 \qquad (5.9)$$

giving the Euler-Lagrange equation:

$$\mu=\frac{\delta T[\rho]}{\delta\rho}+\frac{\delta U[\rho]}{\delta\rho}+\nu(r) \qquad (5.10)$$

The main problem of early theories at this point was the fact that there is no easy expression of T as a functional of ρ, and the minimization of this expression was not possible. The idea of Kohn and Sham was to separate T in a first part, T_s , that represent the kinetic energy of noninteracting particle of density ρ, and the remainder, T_c , that comes from the correlation between particles:

$$T[\rho]=T_s[\rho]+T_c[\rho]$$

in this way, while there is no exactly known functional forms for $T_s[\rho]$ (in the Thomas-Fermi model they used a simple local approximation), it can be easily written in terms of the single-particle orbitals $\phi_i(r)$ of a noninteracting system with density ρ:

$$T_s[\rho]=-\frac{\hbar^2}{2m}\sum_i^N\int dr\phi_i^*(r)\nabla^2\phi(r) \qquad (5.11)$$

This expression is an explicit orbital functional, but an implicit density functional $T_s[\rho]=T_s[\phi_i[\rho]]$. The same separation is possible for the other universal functional, U, leading to the analogue of the well known Coulomb integral of Hartree-Fock, J, that depends on the electrostatic interaction of the charge distribution ρ, plus the remainder U_r:

$$U[\rho] = J[\rho] + U_r[\rho]$$

The sum between the unknown term $T - T_s$ and $U - J$ leads to a new universal functional called the *exchange-correlation functional*:

$$E_{xc}[\rho] = T_c[\rho] + U_r[\rho] = T[\rho] - T_s[\rho] + U[\rho] - J[\rho] \tag{5.12}$$

The corresponding Euler-Lagrange equation of 5.10 takes the form:

$$\mu = \frac{\delta T_s[\rho]}{\delta \rho} + \nu_{KS}(r) \tag{5.13}$$

where

$$\nu_{KS}(r) = \nu_{ext}(r) + \frac{\delta J[\rho]}{\delta \rho} + \frac{\delta E_{xc}[\rho]}{\delta \rho} = \nu_{ext}(r) + \nu_H(r) + \nu_{xc}(r) \tag{5.14}$$

The ground-state density is then calculated by solving a noninteracting-like Schrödinger equation called Kohn-Sham (KS) equation:

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + \nu_{ks}(r) \right] \phi_i(r) = \epsilon_i \phi_i(r) \tag{5.15}$$

and the global wavefunction is then reconstructed as a single Slater determinant made of ϕ_i:

$$\Psi_{KS} = \frac{1}{\sqrt{N!}} |\phi_1(r_1), \phi_2(r_2), ...\phi_N(r_N)| \tag{5.16}$$

It is interesting to stress the analogy between eq 5.15 and eq 5.1, to emphasize the fact that density functional theory replace the problem of solving the complicate N-particle Schrödinger equation by that of minimizing a density functional $E[\rho]$ and then, by use of Kohn Sham equations, replace the problem of minimizing $E[\rho]$ by that of solving a noninteracting Schrödinger equation. Since both ν_H and ν_{HC} depend on ρ, which depends on the ϕ_i , which in turn depend on ν_{KS}, the problem of solving the KS equations is a nonlinear one, just as the Hartree-Fock method. The usual way of solving such problems is to

start with an initial guess for $\rho(r)$, calculate the corresponding $\nu_{KS}(r)$, and then solve the differential eq 5.15 for the ϕ_i. From these a new density can be calculated using 5.4, and the procedure can start again. The process is repeated until it converges in the so-called "self-consistent-field" (SCF) loop. Once the density reached a converged solution ρ_0, the total energy E_0 is easily calculated using this expression:

$$E_0 = \sum_i^N \epsilon_i - \frac{q^2}{2} \int dr \int dr' \frac{\rho_0(r)\rho_0(r')}{|r - r'|} - \int dr \nu_{xc}(r)\rho(r) + E_{xc}[\rho_0] \tag{5.17}$$

5.2 Exchange-correlation functional and its approximations

One of the main criticism of DFT is that, unlike most of the wavefunction based methods, it is not systematically improvable. There is a series of approximations to the exchange-correlation potential that leads to a list of functionals with improved accuracy. Thi DFT analogue to Jacob's ladder was first proposed by Perdew et al.[67] Fig 5.1. The ladder is not based on rigorous mathematical expressions of increasing accuracy, but, despite the lack of such mathematical definitions, we can apply Perdew's ladder to real systems and obtain a hierarchy of results with increasing accuracy.

5.2.1 LDA and LSDA

The local density approximation (LDA) and its extension to fermionic systems local spin density approximation (LSDA), are the first and easiest examples of approximations used in Kohn-Sham DFT. The general idea at their basis is simple: consider a homogeneous electron gas, calculate the exchange-correlation energy per particle e_{xc}, and then obtain the global exchange-correlation energy for a generic system by

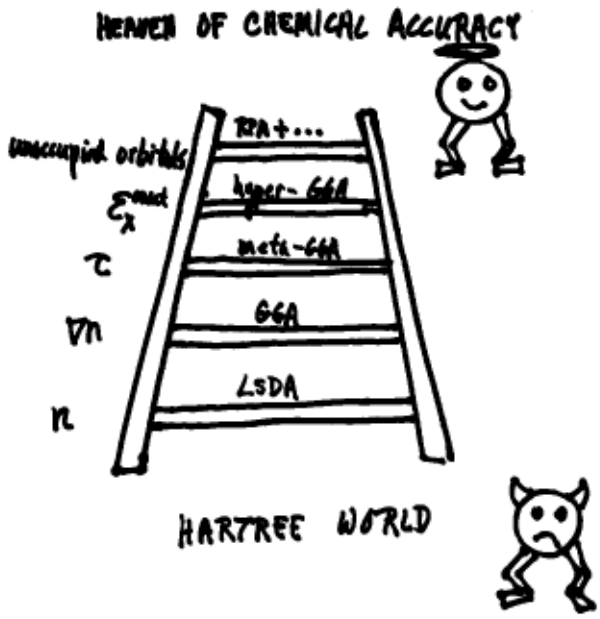


Figure 5.1: Perdew's ladder of density functional approximations

weighting this quantiy by the probability $\rho(r)$, and integrating over all space:

$$E_{xc}^{LDA} = \int dr e_{xc}(\rho(r))\rho(r) \tag{5.18}$$

The exchange-correlation energy per particle, e_{xc}, can be further divided into exchange and correlation contributions:

$$e_{xc}(\rho(r)) = e_x(\rho(r)) + e_c(\rho(r))$$

The exchange part, e_x, for a uniform electron gas is a functional of the density, and can easily be obtained from geometrical consideration:

$$e_x^{LDA} = -\frac{3}{4}\left(\frac{3}{\pi}\rho(r)\right)^{1/3}$$

This expression leads to the well-known $\rho^{4/3}$ dependency of E_x^{LDA}. Analogous analytic expressions for the correlation part e_c are not known, except for the two extreme cases of high- and low-density. All avail-

able correlation functionals use a parametrization of the accurate homogenous electron gas energies obtained with quantum Monte Carlo simulations. The first simulation of homogenous electron gas was performed by Ceperley and Alder, and various mathematical fits to their results are available in modern DFT software:

- Vosko-Wilk-Nusair (VWN)
- Perdew-Zunger (PZ81)
- Cole-Perdew (CP)
- Perdew-Wang (PW92)

5.2.2 GGA

It has been pointed out by many authors, that the major source of error in the local-density approximation is in the exchange energy. This common problem is present since the Thomas-Fermi model and its improvements, include the conventional gradient expansion of von Weizsäcker. In 1986, Perdew[65] proposed a functional that resembles the correct exchange hole, by introducing the reduced gradient variable s:

$$s_\sigma = \frac{|\nabla\rho|}{\rho^{4/3}} \tag{5.19}$$

The per-particle exchange functional is:

$$e_x^{GGA} = -\frac{3}{4}\left(\frac{3}{\pi}\rho(r)\right)^{1/3} F(s) \tag{5.20}$$

and the final form of GGA exchange functional is obtained by integration of 5.20:

$$E_x^{GGA}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int dr \rho^{4/3} F(s) \tag{5.21}$$

whit:

$$F(s) = (c_0 + c_2 s^2 + c_4 s^4 + c_6 s^6)$$
$$c_0 = 1 \qquad c_2 = 1.296 \qquad c_4 = 14 \qquad c_6 = 0.2$$

LSDA can be seen as a zeroth order approximation of this more general method with $c_0 = 1 \qquad c_2 = c_4 = c_6 = 0$. Further improvement to 5.21 have been proposed in the recent years through different expressions of $F(s)$ including some of the most succesful functionals for chemistry, Becke (B88)[7], and solid-state physic Perdew-Burke-Ernzerhof (PBE)[66]. The same scheme has also been applied, with large success, to correlation functionals, leading to a large number of GGA exchange-correlation functionals. The general idea of all GGA functionals, however, is hardly changed from the original formulation: the local-density approximation is improved by including in the functional terms that depend on the gradient of the density, via some function of the reduced density gradient variable s:

$$E_{xc}^{LDA}[\rho] \rightarrow E_{xc}^{GGA}[\rho, \nabla\rho]$$

5.2.3 Hybrid method

If we compare the Hartree-Fock method and density functional theory, we may notice many analogies. These analogies were used in the development of the Hartree-Fock-Kohn-Sham method (HFKS). From the original HFKS method, Becke developed a new kind of exchange functional that includes a fraction of the exact, *nonlocal* exchange, and a complementary fraction of the usual GGA *local* exchange, in the so called Half&Half exchange functional[6]:

$$E_x = \frac{1}{2}E_x^{GGA} + \frac{1}{2}E_x^{HF} \tag{5.22}$$

The exchange-correlation functional might be derived by rigorous analysis of the exchange-correlation hole. Within this rigorous analysis, the

mix of HF exchange and DFT exchange is totally justified by the adiabatic connection, an expression that links the noninteracting case to the full-interacting case by a smooth function of a parameter λ that varies from 0 to 1:

$$E_{xc} = \int_0^1 d\lambda \left\langle \Psi_\rho^\lambda \middle| \hat{V}_{ee} \middle| \Psi_\rho^\lambda \right\rangle - J[\rho] \tag{5.23}$$

Equation 5.22 derives from a simple two-point quadrature of 5.23, evaluating the integrand at $\lambda = 0$ and $\lambda = 1$. At $\lambda = 0$, we simply obtain the non-interacting exchange energy, that is the Hartree-Fock exchange, and at $\lambda = 1$, we have the full exchange+correlation energy. following these arguments, Becke then suggested the introduction of semi-empirical mixing of the HF exchange with the exchange-correlation energy functionals:

$$E_{xc} = E_x^{GGA} + c_x E_x^{HF} + E_c$$

The vast majority of functionals used in modern quatum-chemistry calculations are semi-empiricals hybrids

5.2.4 meta-GGAs, RSHs and DFT-D

The main limitation of GGA functionals, including hybrids, is the failure to treat those effects that require a nonlocal treatment of correlation. Description of dispersion forces, is the main example of failure of conventional DFT. To overcome this limitation, different strategies have been developed in the recent literature. We provide here just a general classification of these new improvements, and we refer to the next section for a more detailed description of the functional forms. Advances beyond GGAs might be divided into:

1. Meta-GGA functionals that depend also on the Kohn-Sham kinetic energy density,τ, the 3rd rung in Perdew's ladder. The most recent examples in this category are the Minnesota family of functionals of Zhao and Truhlar[94;89;95;90;91;92;93] , and the τ-

HCTH[14] family of functionals of Handy and coworkers, and also including the popular BMK[15] functional of Boese and Martin:

$$E_{xc}^{LDA}[\rho] \rightarrow E_{xc}^{GGA}[\rho, \nabla\rho] \rightarrow E_{xc}^{meta-GGA}[\rho, \nabla\rho, \tau] \qquad (5.24)$$

2. Range separated hybrid (RSH) functionals, where the Coulomb operator is separated into long-range and short-range terms, and these two different parts are treated and mixed in appropriate manners. The first to propose this kind of separation was Savin[46;88], and its applications to DFT functionals leads to the LC-BLYP and CAM-B3LYP functionals[46;88], and the recent ωB97 family of functionals of Chai and Head-Gordon[19;20;21].

3. Empirically or semi-empirically corrected functionals, usually referred as DFT-D. The main example here is the B97-D functional of Grimme[33].

4. Double-hybrid functionals that include terms derived from correlated wave-function methods, for instance MP2 perturbation theory, and are considered in the 5th rung in Perdew's ladder. This family includes Grimme's B2-PLYP and mPW2-PLYP[32], and Truhlar's double hybrids[96].

5. Van der Waals density functionals (vdW-DF), propsed by Langreth and coworkers[25]. This functional includes a nonlocal correlation term calculated with an integration over the frequency coordinate of a frequency-dependent density response.

5.3 B97D functional

We turn now the attention to the functionals used in this work.

5.3.1 The general GGA B97 funtionals

This functional is a formulation proposed by Becke in 1997, and is therefore called Becke-97 (B97)[8]. The Becke-97 functional is based on two general ideas. The firse is a remapping of the dimensionless reduced variablw s5.19, to a set of new finite variables $u(s_\sigma^2) \in [0,1]$, with specific forms for the exchange and the correlation:

$$u_{x\sigma}(s_\sigma^2) = \frac{\gamma_{x\sigma} s_\sigma^2}{1 + \gamma_{x\sigma} s_\sigma^2} \tag{5.25}$$

$$u_{c\alpha\beta}(s_{avg}^2) = \frac{\gamma_{c\alpha\beta} s_{avg}^2}{1 + \gamma_{c\alpha\beta} s_{avg}^2} \tag{5.26}$$

$$u_{c\sigma\sigma}(s_\sigma^2 = \frac{\gamma_{c\sigma\sigma} s_\sigma^2}{1 + \gamma_{c\sigma\sigma} s_\sigma^2} \tag{5.27}$$

where $\gamma_{x\sigma} = 0.004$, $\gamma_{c\alpha\beta} = 0.002$, $\gamma_{c\sigma\sigma} = 0.2$ and $s_{avg}^2 = \frac{1}{2}(s_\alpha^2 + s_\beta^2)$. The great advantage of these transformations is that they have linearized the exchange and correlation gradient corrections, and a systematic optimization scheme can be easily created by using a polynomial expansion of the gradient in the new variables u:

$$g = \sum_{i=0}^{M} c_i u^i \tag{5.28}$$

The second idea is to separate the exchange and the correlation into different contributions arising from the interaction of electrons with the same spin, and electrons with opposite spin. The final general-B97 exchange correlation functional is then:

$$E_{xc}^{GGA} = E_{x\alpha\alpha}^{GGA} + E_{x\beta\beta}^{GGA} + c_x E_x^{HF} + E_{c\alpha\alpha}^{GGA} + E_{c\beta\beta}^{GGA} + E_{c\alpha\beta}^{GGA} \tag{5.29}$$

Each term, in (semi-)local form, is expressed as:

$$E_x^{GGA} = \sum_\sigma \int dr e_{x\sigma}^{LSDA}(\rho_\sigma) g_{x\sigma}(s_\sigma^2) \tag{5.30}$$

$$E_{c\alpha\beta}^{GGA} = \int dr e_{c\alpha\beta}^{LSDA}(\rho_\alpha, \rho_\beta) g_{c\alpha\beta}(s_{avg}^2) \tag{5.31}$$

$$E_{c\sigma\sigma}^{GGA} = \int dr e_{c\sigma\sigma}^{LSDA}(\rho_\sigma) g_{c\sigma}(s_\sigma^2) \tag{5.32}$$

where the term "semi-local" describes the fact that the energy density depends only on the electron density and orbitals in an infinitesimal neighborhood of the given position (e_x for the exchange and e_c for the correlation). The global exchange-correlation energy is then obtained by integration of the semi-local term over the entire space $E_{xc} = \int dr e_{xc}$. The optimal coefficients of the expansion series 5.28 are determined by minimization of error functions based on suitable databases of experimental or calculated data. The series is truncated at $M = 2$ (3 terms).

5.3.2 Semiempirically corrected DFT-D for B97

Perhaps motivated by the possibility of a more simplistic approach, several methods for correcting DFT for failures involving noncovalent interactions have involved addition of an empirical correction to the final DFT energy typically of the form $C_6 R^{-6}$, where R represents the interatomic distances and C_6 the dispersion coefficients. B97D functionls follows the Grimme formulation:

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij} \tag{5.33}$$

the damping function is used to scale the dispersion to zero at short distance to avoid near-singularities for small R:

$$f_{dmp}(R_{ij}) = \left[1 + e^{-d\left(\frac{R_{ij}}{R_r} - 1\right)}\right]^{-1} \tag{5.34}$$

where R_r is the sum of atomic vdW radii and they are listed in Grimme work[8] and $d = 20$. C_6^{ij} can be determined via formulas:

$$C_6^{ij} = \sqrt{C_6^i C_6^j}$$

which has been found by Grimme to be consistent, and the functional dependent scale factors, s_6, are determined via parameter fitting. The total energy is given:

$$E_{B97D} = E_{B97} + E_{disp} \tag{5.35}$$

5.4 The range-separated $\omega B97$ functional

In this work we use another functional: ωB97XD, which is a ωB97 functional with dispersion correction (as described in sec 5.3.2. In this section we will describe the ωB97 family. The recentrly proposed ωB97 family of functionals is based on savin's range-separation of the Coulomb operator:

$$\frac{1}{r_{12}} = \frac{erf(\omega r_{12}}{r:12} + \frac{erfc(\omega r_{12}}{r:12} \tag{5.36}$$

The general form of the ωB97 family of functionals, as firstly proposed by Chai and Head-Gordon, is then a long-range corrected (LC-)GGA functional that can be written as follows:

$$E_{xc}^{LC-GGA} = E_x^{SR-GGA} + E_x^{LR-HF} + c_x E_x^{SR-HF} + E_c^{GGA} \tag{5.37}$$

Here, the long- and short-range Hartree-Fock terms are calculated using Savin's Coulomb operator for the two-electron integrals, instead of the regular 1/r 12 Coulomb operator:

$$E_x^{SR-HF} = -\frac{1}{2}\sum_{\sigma}\sum_{i,j}^{occ}\int dr_2 \int dr_1 \Psi_{i\sigma}^*(r_1)\Psi_{j\sigma}^*(r_1) \\ * \frac{erfc(\omega r_{12}}{r_{12}}\Psi_{i\sigma}(r_2)\Psi_{j\sigma}(r_2) \quad (5.38)$$

$$E_x^{LR-HF} = -\frac{1}{2}\sum_{\sigma}\sum_{i,j}^{occ}\int dr_2 \int dr_1 \Psi_{i\sigma}^*(r_1)\Psi_{j\sigma}^*(r_1) \\ * \frac{erf(\omega r_{12}}{r_{12}}\Psi_{i\sigma}(r_2)\Psi_{j\sigma}(r_2) \quad (5.39)$$

$c_x = 0$ for the ωB97 functional, while $c_x \neq 0$ for the ωB97X and ωB97XD. The short-range part of DFT functional is obtained by combination of short-range LSDA exchange and the correction factor $g(s)$ defined for the previous cases:

$$E_x^{SR-GGA} = \sum_{\sigma}\int dr e_{x\sigma}^{SR-LSDA}(\rho_\sigma g_{x\sigma}(s_\sigma^2) \quad (5.40)$$

and the short range LSDA exchange is calculated as an attenuated LSDA function:

$$e_{x\sigma}^{SR-LSDA} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3}\rho_\sigma^{4/3}(r)F(a_\sigma) \quad (5.41)$$

with the attenuation function $F(a_\sigma)$:

$$F(a_\sigma) = 1 - \frac{8}{3}a_\sigma\left[\sqrt{\pi}erf\left(\frac{1}{2a_\sigma}\right) - 3a_\sigma + 4a_\sigma^3 + (2a_\sigma - 4a_\sigma^3)exp\left(\frac{1}{4a_\sigma^2}\right)\right]$$

where:

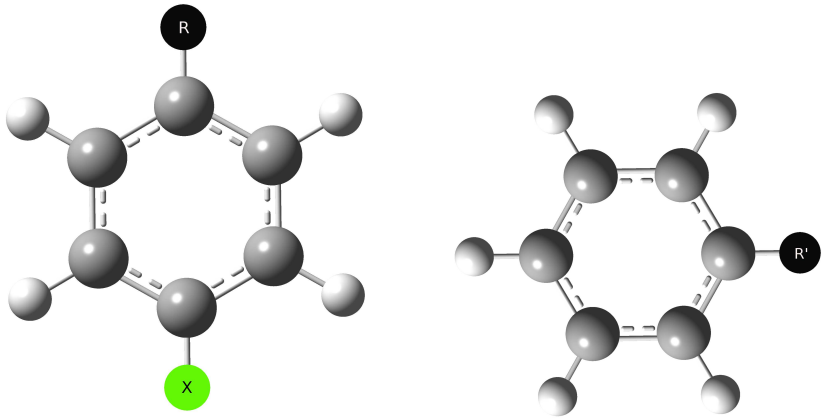
$$a_\sigma = \frac{\omega}{(2k_{F\sigma})} \qquad k_{F\sigma} = (6\pi^2\rho_\sigma(r))^{1/3}$$

Chapter 6

Halogen/π Interactions

6.1 Materials and Methods

As already written, halogen bond is established between an halogn atom and a nucleophilic group due to the presence of a region of positive electrostatic potential, σ-hole. This nucleophilic group can be an atom with lone pairs, for example N,O or S or it can be a system of π electrons of an aromatic ring. This kind of interaction is identified as C-X/π. Non-covalent electrostatic interaction between an halogen and an aromatic ring is of foundamental importance for crystal engineering[11;12;72;13;78;57] like in biological systems. The importance has been noticed for the first time by Bishop et al. in the description of X-ray structures of inclusion compound[11;12;72;13]. Complex halogen/π complexes are intermediates in reaction of electrophilic halogenation of alkenes, alkynes and aromatic systems[2;3;22]. In this work we studied the effect of substitution on a system where the halogen bond is established between an halobenzene (Figure 6.1 (a)) and a benzene (Figure 6.1 (b)).



(a) $X = Br, I$ $R = H, NO_2, CN$ (b) $R' = H, NH_2, OH, CH_3$

Figure 6.1: The two interacting molecules

For each complex we perform DFT calculation using GAUSSIAN[29] starting from different geometries for searching the Potential Energy Surface*.

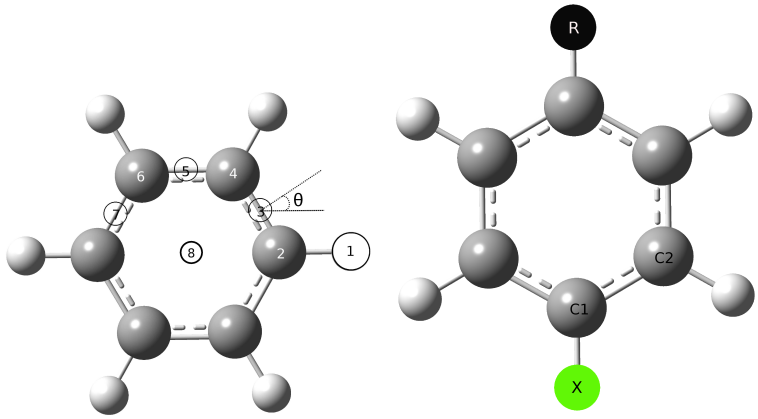
X	R	R'
	H	CN
	H	NO2
Br	CH3	H
	OH	H
	NH2	H
	H	CN
	H	NO2
I	CH3	H
	OH	H
	NH2	H

Table 6.1: Complex

Table 6.1 show all complex. For each of this complex we searched the PES starting from different geometries. Each geometry is definied by 2 parameters: position (P) and angle (θ). Position is the interception between C-X axis and the ring plane of the substituted benzene with

*From now on we refer to it as PES

angle 90°,Figure 6.2 (a) shows all the position.

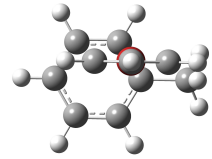


(a) (b)

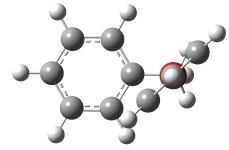
Figure 6.2: Parameters

Angle is the dihedral formed by atoms 1-X-C1-C2 (for numeration see Figure 6.2).

P	θ	P	θ
	0		0
1	45	5	45
	90		90
	0		0
2	45	6	45
	90		90
	0		0
3	45	7	45
	90		90
	0		0
4	45	8	30
	90		60
			90



(a) P=3 ; Θ=0



(b) P=1 ; Θ=45

Figure 6.3: All geomtries

Table 6.3 shows all the starting geometries[†] and Figure 6.3 (a) and (b) shows two examples of this geometries. For cases where halobenzene is substituted we have less geometries because some of them are equivalent.

P	θ	P	θ	P	θ	P	θ
	0		0		0		0
1	45	2	45	3	45	8	30
	90		90		90		60
							90

Table 6.2: Geometries for Halobenzene substituted

Each geometries of table 6.2 and table 6.3 were submitted to geometry otpimization at the B97D/aug-cc-PVDZ level to searching the PES. For each stationary point the interaction energy (ΔE_{int}) was calculated. The basis set superposition error (BSSE) was taken into acount by means of Boys-Bernardi counterpoise method[17]. Each stationary points were classificated in Transition State or Minima[‡]. Cluster analysis were performed to identify the unique stationary points both on TS and min. This unique TS and min were submitted to further geometry optimization at the WB97XD/aug-cc-pvdz level to refine results.

[†]In this work we refer to one geometry as R_R'_P_θ. For example: h_ch3_1_45

[‡]From now on we refer to Transition State as TS and Minima as min

6.2 Bromine

CH3_H				
	B97D		WB97XD	
1_45	Min	-5.19	Min	-6.10
7_90	Min	-2.58	TS	-2.51
T_30	Min	-2.54	TS	-2.48
T_60	Min	-2.56	Min	-2.49
4_90	TS	-2.60	TS	-2.51
7_0	TS	-2.54	TS	-2.49
T_0	TS	-2.56	TS	-2.51
T_90	TS	-2.54	TS	-2.51

OH_H				
	B97D		WB97XD	
1_45	Min	-5.92	Min	-6.72
1_90	Min	-5.92	Min	-6.72
6_90	Min	-2.19	Min	-2.17
7_90	Min	-5.80	Min	-6.67
T_90	Min	-2.33	Min	-2.26
2_90	TS	-2.54	Min	-2.26

NH2_H				
	B97D		WB97XD	
1_45	Min	-4.23	Min	-6.09
2_45	Min	-2.17	TS	-2.12
4_90	Min	-2.16	Min	-2.08
5_90	Min	-2.16	TS	-2.14
6_45	Min	-2.19	Min	-2.14
7_45	Min	-3.92	TS	-4.90

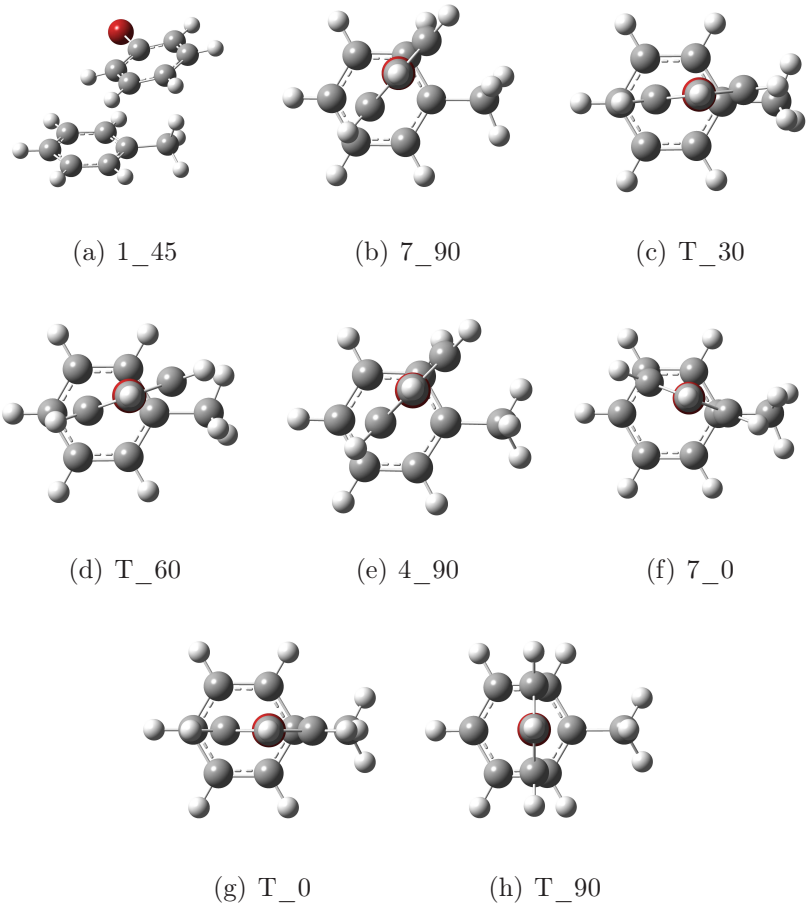
H_CN				
	B97D		WB97XD	
2_45	Min	-2.46	Min	-2.49
T_30	TS	-2.32	TS	-2.46
T_0	TS	-2.32	TS	-2.46
3_45	TS	-2.46	Min	-2.50
2_90	TS	-2.41	TS	-2.46

H_NO2				
	B97D		WB97XD	
3_90	Min	-2.55	Min	-2.56
T_30	TS	-2.41	TS	-2.52
T_0	TS	-2.41	TS	-2.52
3_45	TS	-2.55	Min	-2.55
2_90	TS	-2.50	TS	-2.53
2_0	TS	-2.56	Min	-2.55

Table 6.3: Unique stationary points for bromine

Table 6.3 shows the interaction energies for all unique stationary points founded both on B97D level and WB97XD. In the Tableis showed also the type of stationary point founded based on the value of the first frequency value.

6.2.1 CH3_H

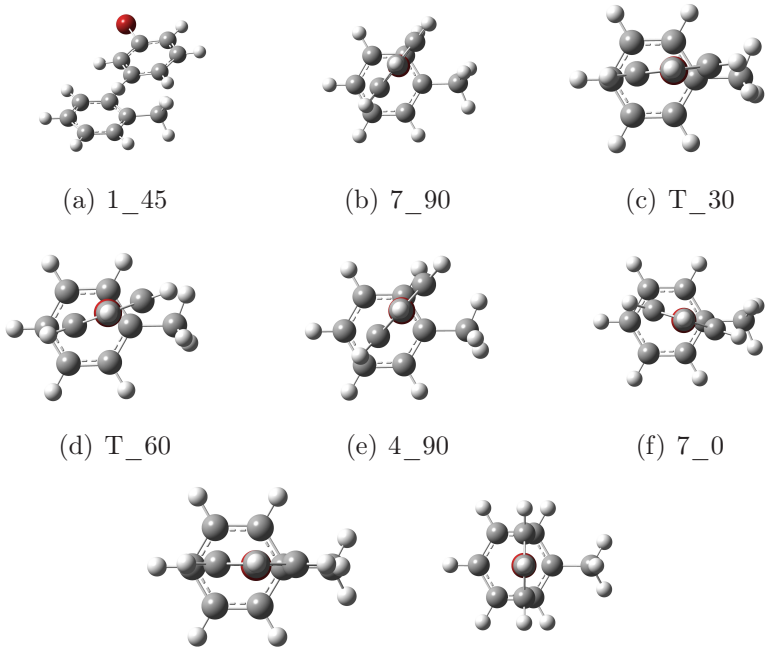


(a) 1_45 (b) 7_90 (c) T_30

(d) T_60 (e) 4_90 (f) 7_0

(g) T_0 (h) T_90

Figure 6.4: Unique stationary point at B97D/aug-cc-pvdz level



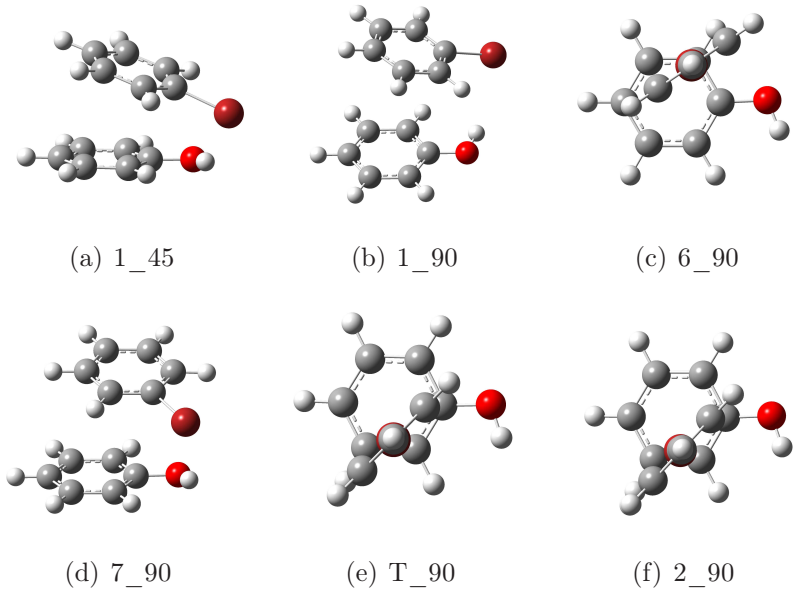
(a) 1_45 (b) 7_90 (c) T_30

(d) T_60 (e) 4_90 (f) 7_0

(g) T_0 (h) T_90

Figure 6.5: Unique stationary point at WB97XD/aug-cc-pvdz level

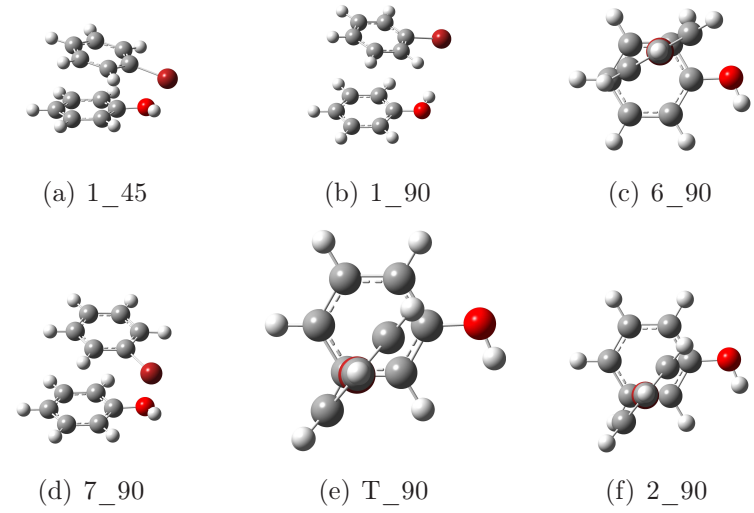
6.2.2 OH_H



(a) 1_45 (b) 1_90 (c) 6_90

(d) 7_90 (e) T_90 (f) 2_90

Figure 6.6: Unique stationary point at B97D/aug-cc-pvdz level

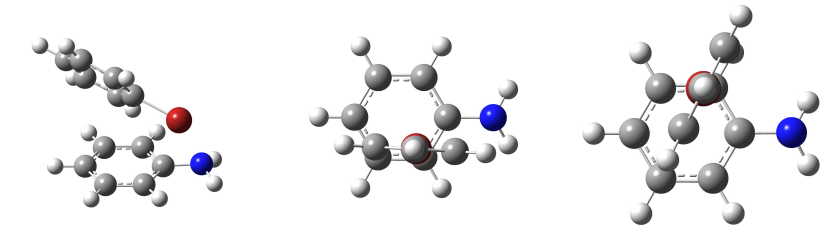


(a) 1_45 (b) 1_90 (c) 6_90

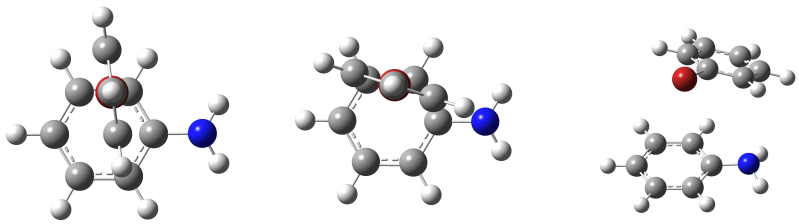
(d) 7_90 (e) T_90 (f) 2_90

Figure 6.7: Unique stationary point at WB97XD/aug-cc-pvdz level

6.2.3 NH2_H

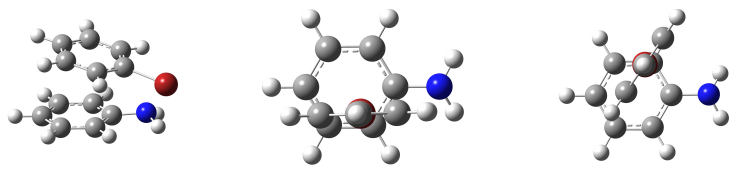


(a) 1_45 (b) 2_45 (c) 4_90

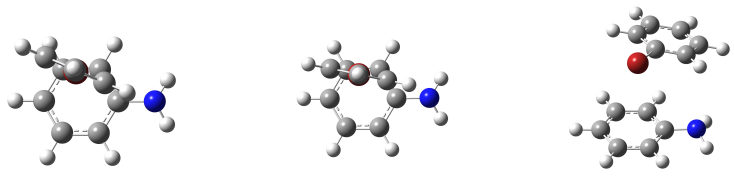


(d) 5_90 (e) 6_45 (f) 7_45

Figure 6.8: Unique stationary point at B97D/aug-cc-pvdz level



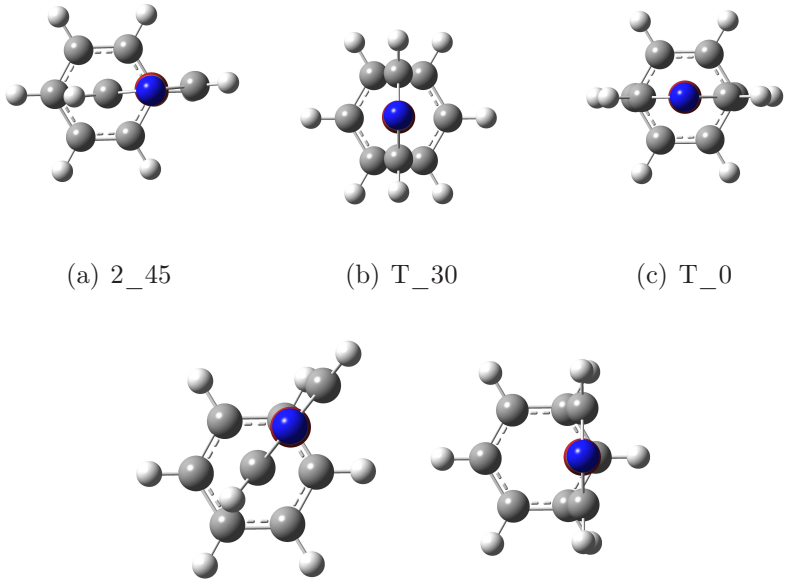
(a) 1_45 (b) 2_45 (c) 4_90



(d) 5_90 (e) 6_45 (f) 7_45

Figure 6.9: Unique stationary point at WB97XD/aug-cc-pvdz level

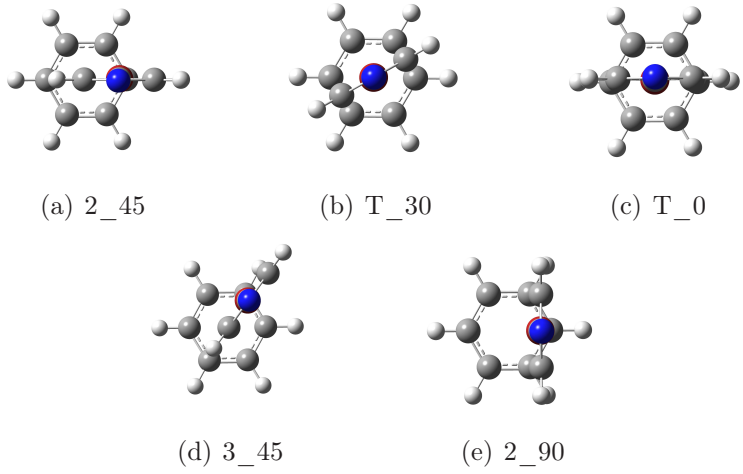
6.2.4 H_CN



(a) 2_45 (b) T_30 (c) T_0

(d) 3_45 (e) 2_90

Figure 6.10: Unique stationary point at B97D/aug-cc-pvdz level

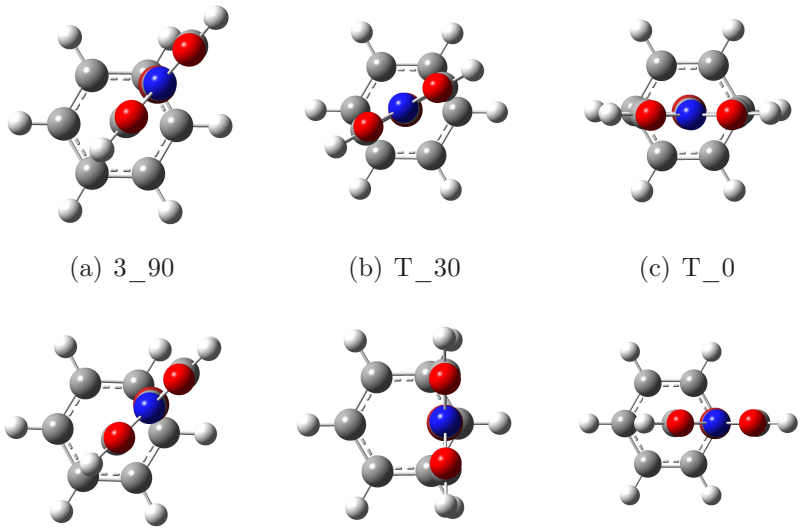


(a) 2_45 (b) T_30 (c) T_0

(d) 3_45 (e) 2_90

Figure 6.11: Unique stationary point at WB97XD/aug-cc-pvdz level

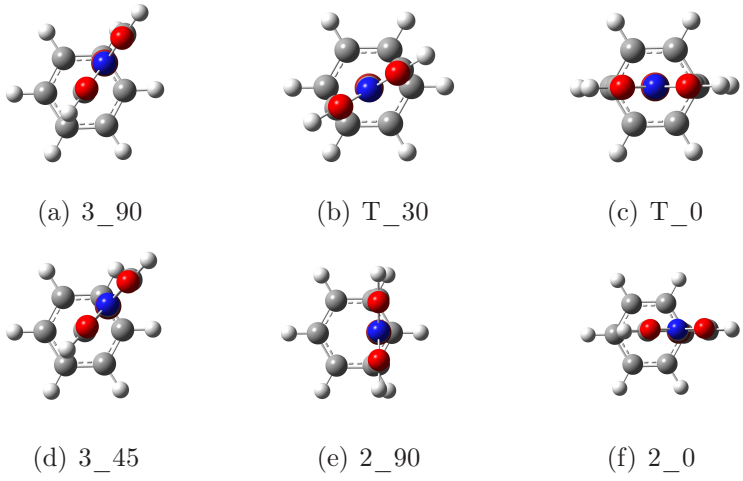
6.2.5 H_NO2



(a) 3_90 (b) T_30 (c) T_0

(d) 3_45 (e) 2_90 (f) 2_0

Figure 6.12: Unique stationary point at B97D/aug-cc-pvdz level



(a) 3_90 (b) T_30 (c) T_0

(d) 3_45 (e) 2_90 (f) 2_0

Figure 6.13: Unique stationary point at WB97XD/aug-cc-pvdz level

6.2.6 Result

As we can see not all the optimization identify a stationary point with an halogen bond interaction. Usually stationary points where we have halogen bond interaction have an average interaction energy lower than other structure. This is probabily due to stacking between the two aromatic rings. In this work we focused our attention on the halogen bond interaction wichi is highly directional, so we don't study this kind of stationary points.

Substituent	Place	ΔG_{int}	Type
CH3	R'	-2.56	Min
CH3	R'	-2.56	TS
NH2	R'	-2.17	Min
OH	R'	-2.26	Min
OH	R'	-2.54	TS
CN	R	-2.46	Min
CN	R	-2.38	TS
NO2	R	-2.55	Min
NO2	R	-2.49	TS

The Table 6.2.6 shows a syntesis of the substituents and the interactions energies. The trend between force of interaction for R substituents is NO2 $>$ CN. For R' substituents is CH3 $>$ OH $>$ NH2.

6.3 Iodine

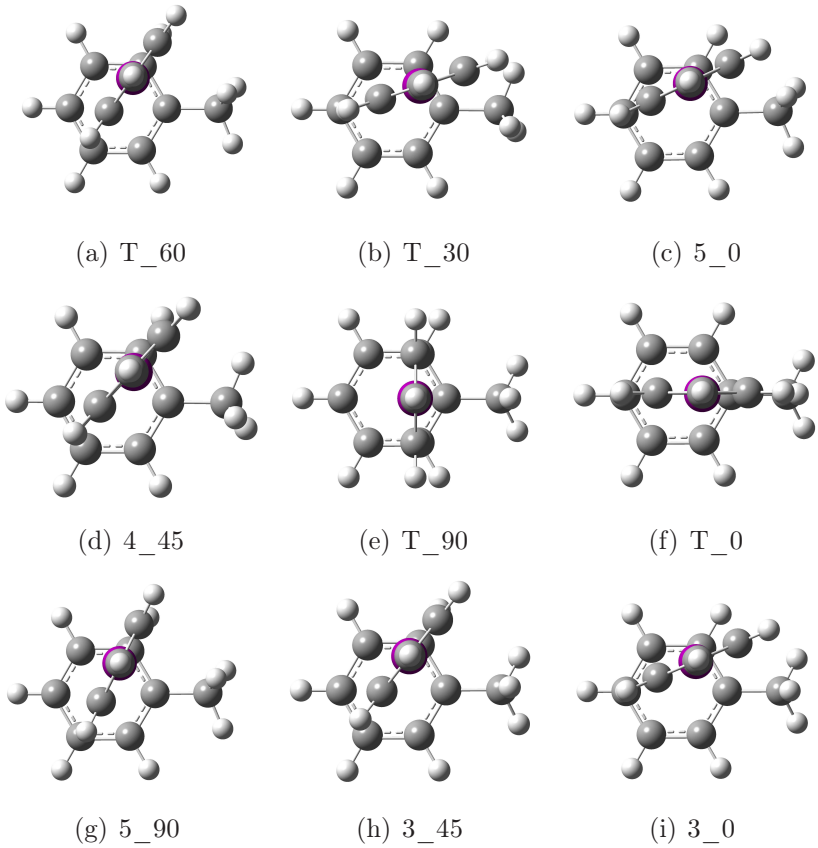
CH3_H					NH2_H				
	B97D		WB97XD			B97D		WB97XD	
T_60	Min	-4.07	TS	-3.69	T_60	Min	-3,71	TS	-3.33
T_30	Min	-4.03	Min	-3.68	T_30	Min	-3,72	Min	-3.34
5_0	Min	-4.06	Min	-3.68	7_90	Min	-3,65	Min	-3.33
4_45	Min	-4.04	TS	-3.69	6_45	Min	-3,70	Min	-3.31
T_90	TS	-4.00	Min	-3.67	1_45	Min	-3,86	Min	-3.91
T_0	TS	-4.02	TS	-3.66	3_90	Min	-3,70	TS	-3.30
5_90	TS	-4.07	TS	-3.69	5_90	TS	-3,70	TS	-3.30
3_45	TS	-4.09	TS	-3.69	3_45	TS	-3,72	Min	-3.34
3_0	TS	-4.08	Min	-3.68					

OH_H					H_CN				
	B97D		WB97XD			B97D		WB97XD	
T_90	Min	-3,68	Min	-3.32	T_30	Min	-3,87	Min	-3.60
6_90	Min	-3,56	Min	-3.25	2_0	TS	-3,87	TS	-3.60
4_90	Min	-3,56	Min	-3.24	2_90	TS	-3,87	TS	-3.58
7_0	TS	-3,55	Min	-3.30	T_0	TS	-3,68	TS	-3.54

H_NO2				
	B97D		WB97XD	
3_90	Min	-3,99	Min	-3.66
2_90	TS	-3,94	TS	-3.65
T_0	TS	-3,79	TS	-3.61
T_30	TS	-3,79	TS	-3.61

Table 6.4: Unique stationary points for iodine

6.3.1 CH3_H

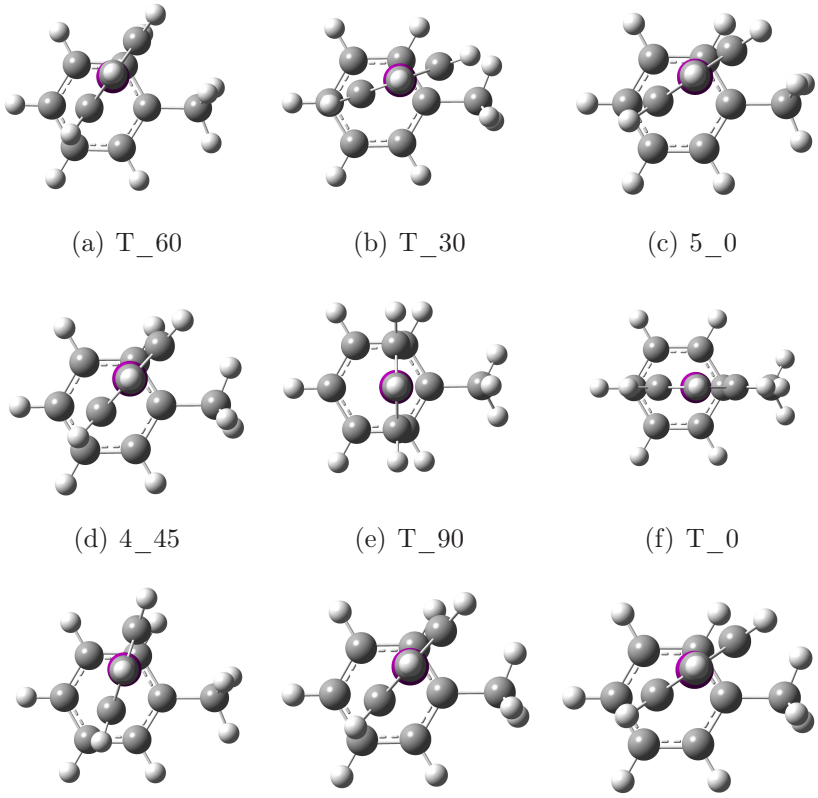


(a) T_60 (b) T_30 (c) 5_0

(d) 4_45 (e) T_90 (f) T_0

(g) 5_90 (h) 3_45 (i) 3_0

Figure 6.14: Unique stationary point at B97D/aug-cc-pvdz level



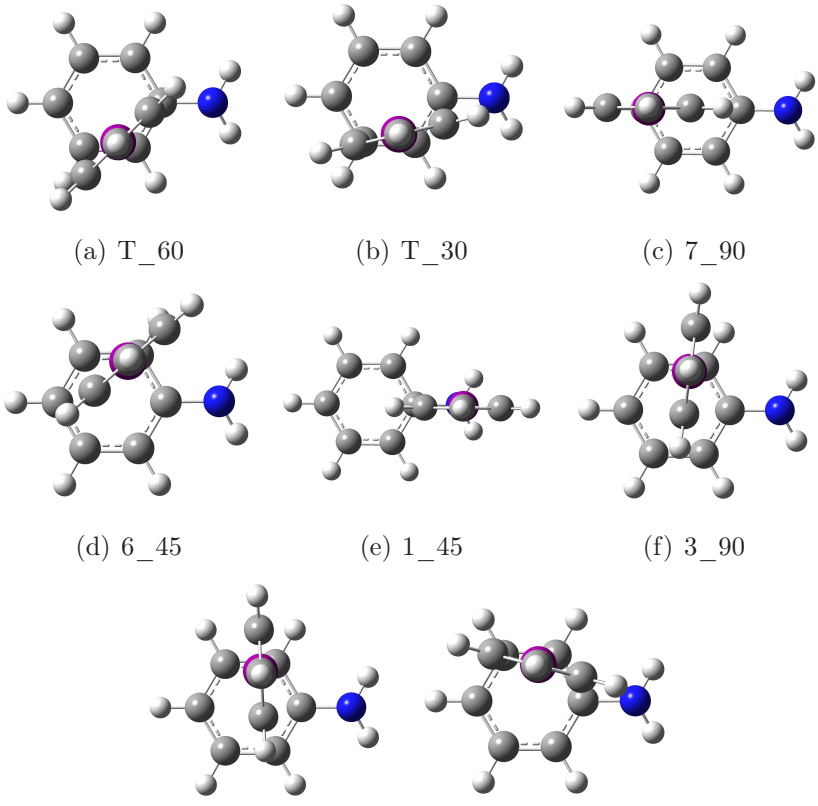
(a) T_60 (b) T_30 (c) 5_0

(d) 4_45 (e) T_90 (f) T_0

(g) 5_90 (h) 3_45 (i) 3_0

Figure 6.15: Unique stationary point at WB97XD/aug-cc-pvdz level

6.3.2 NH2_H

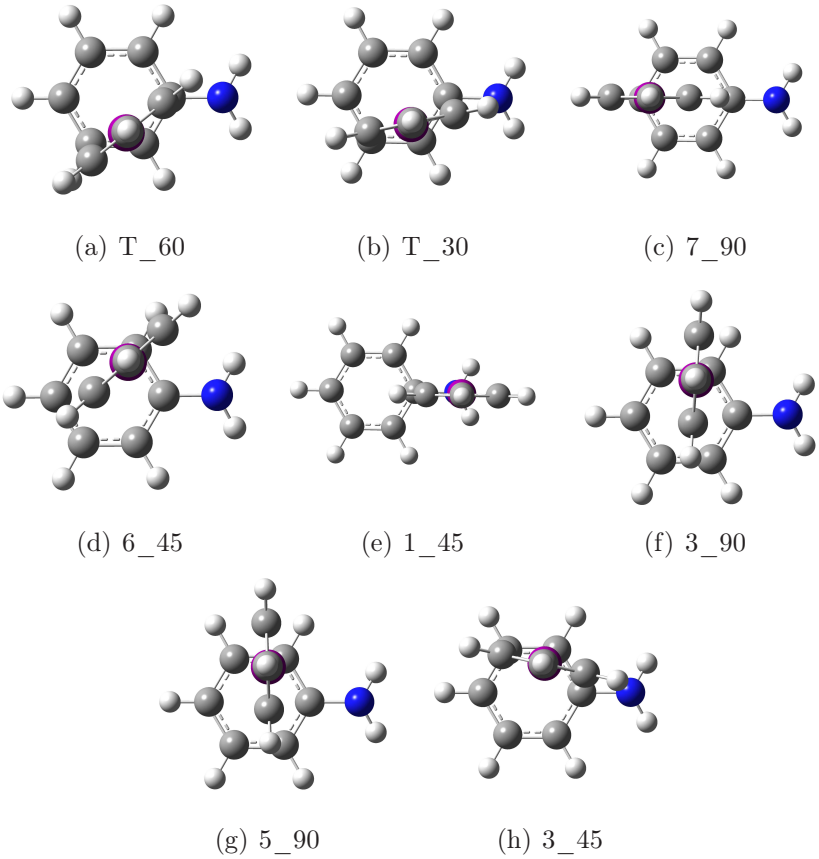


(a) T_60 (b) T_30 (c) 7_90

(d) 6_45 (e) 1_45 (f) 3_90

(g) 5_90 (h) 3_45

Figure 6.16: Unique stationary point at B97D/aug-cc-pvdz level



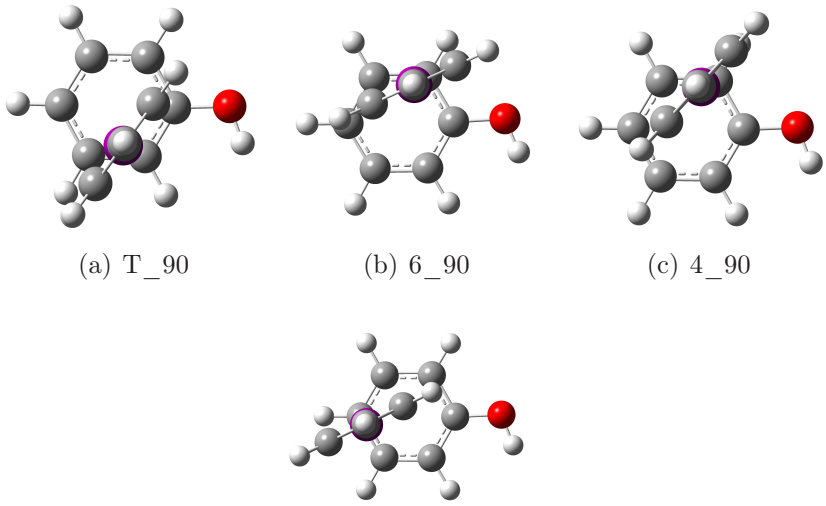
(a) T_60 (b) T_30 (c) 7_90

(d) 6_45 (e) 1_45 (f) 3_90

(g) 5_90 (h) 3_45

Figure 6.17: Unique stationary point at WB97XD/aug-cc-pvdz level

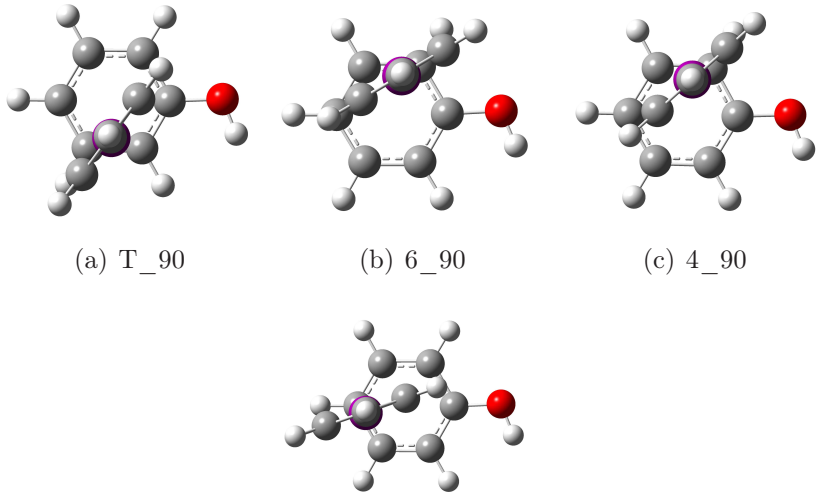
6.3.3 OH_H



(a) T_90 (b) 6_90 (c) 4_90

(d) 7_0

Figure 6.18: Unique stationary point at B97D/aug-cc-pvdz level

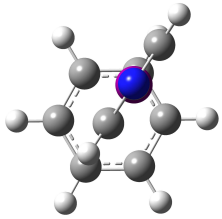


(a) T_90 (b) 6_90 (c) 4_90

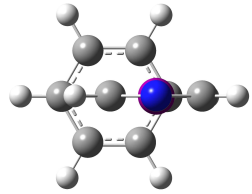
(d) 7_0

Figure 6.19: Unique stationary point at WB97XD/aug-cc-pvdz level

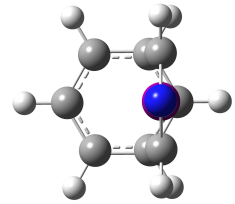
6.3.4 H_CN



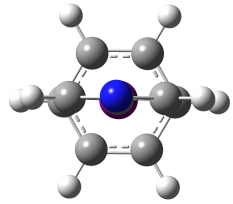
(a) T_30



(b) 2_0

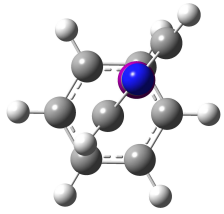


(c) 2_90

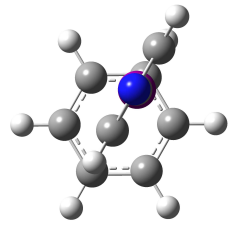


(d) T_0

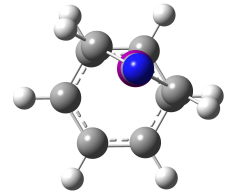
Figure 6.20: Unique stationary point at B97D/aug-cc-pvdz level



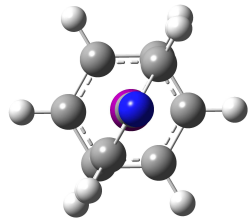
(a) T_30



(b) 2_0



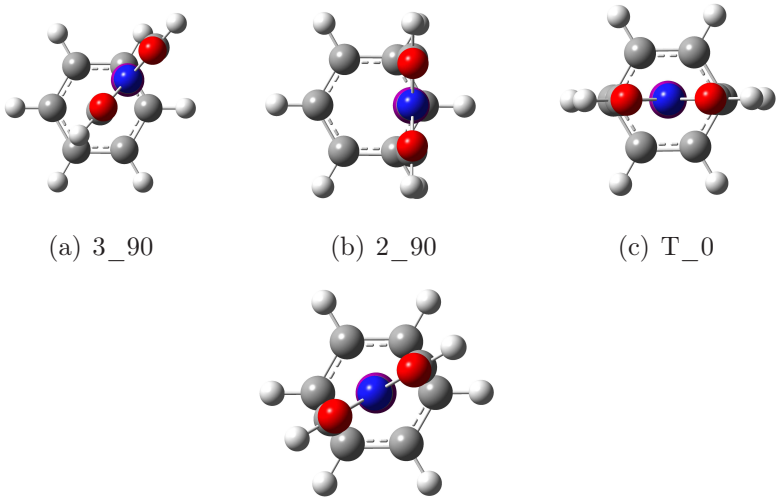
(c) 2_90



(d) T_0

Figure 6.21: Unique stationary point at WB97XD/aug-cc-pvdz level

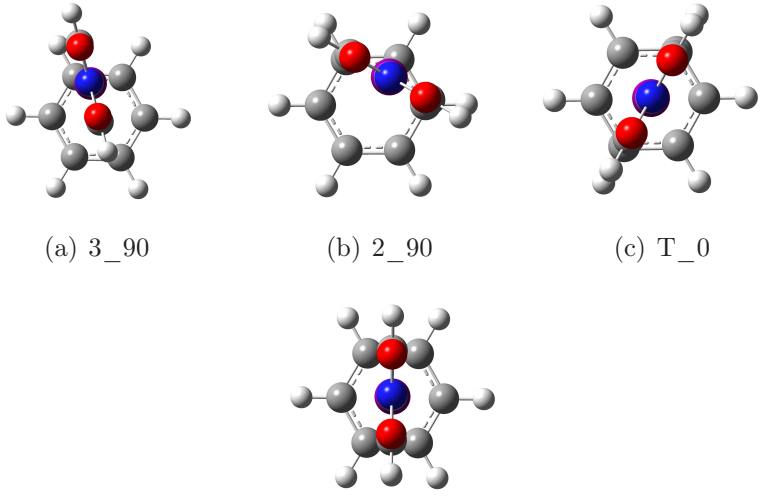
6.3.5 H_NO2



(a) 3_90 (b) 2_90 (c) T_0

(d) T_30

Figure 6.22: Unique stationary point at B97D/aug-cc-pvdz level



(a) 3_90 (b) 2_90 (c) T_0

(d) T_30

Figure 6.23: Unique stationary point at WB97XD/aug-cc-pvdz level

6.3.6 Result

For Iodine all the stationary points identify an halogen bond between the two molecules. Force of interactions grows up with the dimension of halogen. Iodine is the halogen which form the strongest interaction.

Substituent	Place	ΔG_{int}	Type
CH3	R'	-3.69	Min
CH3	R'	-3.68	TS
NH2	R'	-3.72	Min
NH2	R'	-3.71	TS
OH	R'	-3.27	Min
OH	R'	-3.30	TS
CN	R	-3.87	Min
CN	R	-3.81	TS
NO2	R	-3.99	Min
NO2	R	-3.84	TS

The trend between force of interaction for R substituents is NO2 > CN. For R' substituents is NH2 > CH3 > OH.

6.4 Discussion

As we show in previous sections we identify the best substituents for halobenzene to reinforce the strenght of the halogen bond interaction. Comparing the strenght of interaction between halogen we found that in general Iodine form a stronger interaction. Iodine, in fact, form always an interaction stronger than bromine of at least 45% for every substituent. Table 6.4 shows a comparison between interaction energies of both halogen.

Substituent	Iodine	Bromine
CH3	-4.05	-2.56
OH	-3.72	-2.17
NH2	-3.27	-2.26
CN	-3.87	-2.46
NO2	-3.99	-2.55

When we perform a comparison between substituents we can identify a trend for substitution on halobenzene and benzene. For R substitution [§] the trend both for Bromine and Iodine is: NO2 > CN. Taking into account the Hammett the trend between electron withdrawal substituent (as NO2 and CN) is NO2 > CN. This results confirm the experimental trend. For R' substitution we found two different trends. For Bromine: CH3 > OH > NH2; for Iodine: NH2 > CH3 > OH. Taking into account the Hammett constant the theoretical trend is: NH2 > OH > CH3. In this case we don't reproduce the expected trend. In this study we can identify the best R and R' substituent both for iodine and bromine. The best R substituent is NO2 for both halogen, the best R' substituent is CH3 for bromine and NH2 for Iodine. If we take a look to the Table 6.4 and Table 6.3 we can see that when we perform optimization on the stationary points at B97D/level with WB97XD functional the minima became a trasition state. This behavior has not been explained, further studies are needed on the potential surface.

[§]R is substitution on halobenzene and R' on benzene

Bibliography

[1] L. Amos and A. Klung. Arrangement of subunits in flagellar microtubules. *Journal of Cell Science*, (14), 1974.

[2] C. Angew and A. Legon. *Chem.Int.Ed.*, (39), 1999.

[3] C. Angew and D. Lenoir. *Chem.Int.Ed.*, (42), 2003.

[4] A.R. A.R. Voth, F.A. Hays, and P.S. Ho. Directing macromolecular conformation through halogen bonds. *Proc.Natl.Acad.Sci.*, (104), 2007.

[5] P. Auffinger, F.A. Hays, E. Westhof, and P.S. Ho. Halogen bonds in biological molecules. *Proc.Natl.Acad.Sci.*, (101), 2004.

[6] A. D. Becke. *J Chem Phys*, (98), 1993.

[7] A.D. Becke. *Phys rev A*, (38), 1988.

[8] A.D. Becke. Density-functional thermochemistry. v. systematic optimization of exchange-correlation functionals. *J.Chem.Phys.*, (107), 1997.

[9] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Research*, (28), 2000.

[10] D. L. Beveridge and F. M. DiCapua. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry*, (18), 1989.

[11] R. Bishop, D.C. Craig, M.L. Scudder, and A.N.M.M. Rahman. *Cryst. Eng. Commun.*, (4), 2002.

[12] R. Bishop, D.C. Craig, M.L. Scudder, and A.N.M.M. Rahman. *Cryst.Eng.Commun.*, (5), 2003.

[13] R. Bishop, D.C. Craig, M.L. Scudder, S.F. Alshahateet, and A.N.M.M. Rahman. *Mol.Cryst. Liq*, 2005.

[14] A. Boese and N.C. Handy. *J Chem Phys*, (116), 2002.

[15] A.D. Boese and J.M.L. Martin. *J Chem Phys*, (121), 2004.

[16] D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, and R. C. Wade. *J Med Chem*, (32), 1989.

[17] S.F. Boys and F. Bernardi. *Mol.Phys.*, (19), 1970.

[18] E. Cariati, A. Forni, S. Biella, P. Metrangolo, F. Meyer, G. Resnati, S. Righetto, E. Tordin, and R. Ugo. Tuning second order nlo responses through halogen bonding. *Chem.Commun*, (25), 2007.

[19] J.D. Chai and M Head-Gordon. *J Chem Phys*, (128), 2008.

[20] J.D. Chai and M. Head-Gordon. *Phys. Chem. Chem. Phys*, (10), 2008.

[21] J.D. Chai and M. Head-Gordon. *J Chem Phys*, (131), 2009.

[22] C. Chiappe and D. Lenoir. *Chem.Eur.J.*, (9), 2003.

[23] B. Cunningham and J. Wells. High-resolution epitope mapping of hgh-receptor interactions by alanine-scanning mutagenesis. *Science*, (244), 1989.

[24] W. L. DeLano. Unraveling hot spots in binding interfaces: progress and challenges. *Current opinion in structural biology*, (12), 2002.

[25] M. Dion, H. Rydberg, E. Schroeder, D.C. Langreth, and B.I. Lundqvist. *Phys. Rev. Let.*, (92), 2004.

[26] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee. *J Comput Aided Mol Des*, (11), 1997.

[27] D. Frenkel and B. Smit. Understanding molecular simulation: from algorithms to applications. *Academic Press*, 2002.

[28] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin. *J Med Chem*, (47), 2004.

[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

[30] H. Gohlke, M. Hendlich, and G. Klebe. *J Mol Biol*, (295), 2000.

[31] P. J. Goodford. *J Med Chem*, (28), 1985.

[32] S. Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J Chem Phys*, (124), 2006.

[33] S. Grimme. *J. Chem. Phys*, (131), 2009.

[34] L.A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, J.M. Plancher, G Hartmann, D.W. Banner, W. Haap, and F. Dietrich. Systematic investigation of halogen bonding in protein–ligand interactions. *Angew.Chem.Int.*, (50), 2011.

[35] M. Hendlich, A. Bergner, J. Gunther, and G Klebe. *Journal of Molecular Biology*, 326(2), 2003.

[36] S. Huo, I. Massova, , and P. A. Kollman. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *Journal of Computational Chemistry*, (23), 2002.

[37] T. Imakubo, H. Sawa, and R. Kato. Novel radical cation salts of organic π-donors containing iodine atom(s): the first application of strong intermolecular-i$\cdots$x-(x = cn, halogen atom) interaction to molecular conductors. *Synth.Met.*, (73), 1995.

[38] T. Imakubo, A. Miyake, H. Sawa, and R. Kato. Synthesis and physical properties of (diets) 2 [au (cn) 4]: A new θ-salt with a unique donor... anion network. *Synth.Met.*, (120), 2001.

[39] A.N. Jain. *J Med Chem*, (46), 2003.

[40] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. *J Mol Biol*, (267), 1997.

[41] M. Jordan and L. Wilson. Microtubules as target for anticancer drugs. *Nature Reviews Cancer*, (4), 2004.

[42] W.L. Jorgensen, J. Chandrasekhar, J. Madura, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J.Chem.Phys*, (79), 1983.

[43] R Kato, T. Imakubo, H. Yamamoto, R. Maeda, M. Fujiwara, and H. Sawa. An application of supramolecular chemistry to molecular conductors. *Cryst.Liq.Cryst.*, (380), 2002.

[44] S. Kumar and R. Nussinov. Close-range electrostatic interactions in proteins. *ChemBioChem*, 7(3), 2002.

[45] A.C. Legon. The halogen bond: an interim perspective. *Phys.Chem.Phys*, (12), 2010.

[46] T. Leininger, H. Stoll, H. Werner, and A. Savin. *Chemical Physics Letters*, (275), 1997.

[47] J.P.M. Lommerse, A.J. Stone, R. Taylor, and F.H. Allen. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J.Am.Chem.Soc*, (118), 1996.

[48] J. Lowe, H. Li, K.H. Downing, and E. Nogales. Refined structure of $\alpha\beta$-tubulin at 3.5 a resolution. *J.Mol.Biol*, (313), 2001.

[49] Y. Lu, Y. Wang, and W. Zhu. Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. *Phys.Chem.Chem.Phys*, (12), 2010.

[50] I. Massova and P. A. Kollman. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society.*, 1999.

[51] I. Massova and P.A. Kollman. Combined molecular mechanical and continuum solvent approach (mm-pbsa/gbsa) to predict ligabd binding. *Perspectives in Drug Design*, (18), 2000.

[52] E. L. Mehler and T. Solmajer. *Protein Eng*, (4), 1991.

[53] E. C. Meng, B. K. Shoichet, and I. D. Kuntz. *J Comput Chem*, (13), 1992.

[54] P. Metrangolo and G. Resnati. Halogen bonding: a paradigm in supramolecular chemistry. *Chem.Eur.J.*, (7), 2001.

[55] P. Metrangolo and G. Resnati. *Hologen bonding: funamentals and applications*. Springer, 2008.

[56] P. Metrangolo and G. Resnati. Chemistry. halogen versus hydrogen. *Science*, (321), 2008.

[57] I. Mizutani, Y. Ono, K. Kishikawa, I. Azumaya, K. Yanaguchi, S. Kohmoto, and H. Mash. *Cryst. Growth. Des.*, (6), 2006.

[58] I. S. Moreira, P. A. Fernandes, , and M. J. Ramos. Computational alanine scanning mutagenesis - an improved methodological approach. *Journal of Computational Chemistry*, (28), 2006.

[59] I. S. Moreira, P. A. Fernandes, , and M. J. Ramos. Unraveling the importance of protein-protein interaction: application of a computational alanine-scanning mutagenesis to the study of the igg1 streptococcal protein g (c2 fragment) complex. *The Journal of Physical Chemistry. B*, (110), 2006.

[60] G.M Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*, (19), 1998.

[61] G.M Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, and A.J. Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J Comput Chem*, (30), 2009.

[62] E. Nogales and H.W. Wang. Structural mechanisms underlying nucleotide-dependent self-assembly of tubulin and its relatives. *Current Opinion in Structural Biology*, (16), 2006.

[63] E. Nogales, M. Whittaker, A.R. Milligan, and K.H Downing. High-resolution model of the microtubule. *Cell*, (96), 1999.

[64] E. Parisini, P. Metrangolo, T. Pilati, G. Resnati, and G. Terraneo. Halogen bonding in halocarbon–protein complexes: a structural survey. *Chem.Soc.Rev.*, (40), 2011.

[65] J.P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, (33), 1986.

[66] J.P. Perdew, K. Burke, and M. Ernzerhof. *Phys. rev.Lett.*, (77), 1996.

[67] J.P. Perdew, A. Ruzsinszky, L.A. Constantin, J. Sun, and G.I. Csonka. Some fundamental issues in ground-state density functional theory: A guide for the perplexed. *Journal of Chemical Theory and Computation*, (5), 2009.

[68] C. Perez and A. R. Ortiz. *J Med Chem*, (44), 2001.

[69] S. Pieraccini, G. Saladino, G. Cappelletti, D. Cartelli, P. Francescato, G. Speranza, P. Manitto, and M. Sironi. In silico design of tubulin-targeted antimitotic peptides. *Nature Chemistry*, (1), 2009.

[70] P. Politzer, P. Lane, M.C. Concha, Y. Ma, and S. Murray. Halogen bonding: the σ-hole. *J.Mol.Model*, (13), 2007.

[71] P. Politzer, J.S. Murray, and Clark T. Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys.Chem.Chem.Phys.*, (12), 2010.

[72] R. R. Bishop, D.C. Craig, M.L. Scudder, and A.N.M.M. Rahman. *Org. Biomol. Chem.*, (2), 2004.

[73] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. *J Mol Biol*, (261), 1996.

[74] S Rendine, S. Pieraccini, A. Forni, and M Sironi. Halogen bonding in ligand–receptor systems in the framework of classical force fields. *Phys.Chem.Chem.Phys.*, (13), 2011.

[75] K.E. Riley, J.S. Murray, P. Politzer, M.C. Concha, and P. Hobza. Br $\cdots$ O complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone. *J.Chem.Theory Comput.*, (5), 2009.

[76] J.P. Ryckaert, G. Cicciotti, and H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J.Comput.Phys*, (23), 1977.

[77] H.T. Schek, M.K. Gardner, J. Cheng, D.J. Odde, and A.J. Hunt. Microtubule assembly dynamics at the nanoscale. *Current Biology*, (17), 2007.

[78] A. Shterenberg, M. Kapon, K. Suwinska, Y. Eichen, and B. Turner. *Chem. Commun*, 2001.

[79] P. F. W. Stouten, C. Frommel, H. Nakamura, and C. Sander. *Mol Simul*, (10), 1993.

[80] T. P. Straatsma and J. A. McCammon. Multiconfiguration thermodynamic integration. *The Journal of Chemical Physics*, (95), 1991.

[81] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting free energy calculations: a theoretical connection to mm/pbsa and direct calculation of the association free energy. *Biophysical journal*, (86), 2004.

[82] H. Teramoto, R.and Fukunishi. *J Chem Inf Model*, (48), 2008.

[83] R.D. Vale, C.M. Coppin, F. Malik, F.J. Kull, and R.A. Milligan. Tubulin gtp hydrolysis influences the structure, mechanical properties, and kinesin-driven transport of microtubules. *Journal of Biological Chemistry*, (269), 1994.

[84] C. M. Venkatachalam, X. Jiang, T. Oldfield, and M. Waldman. *J Mol Graph Model*, (21), 2003.

[85] A.R. Voth, P. Khuu, K. Oishi, and P.S. Ho. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat.Chem.*, (1), 2009.

[86] R. Wang, Y. Lu, and S. Wang. *J Med Chem*, (46), 2003.

[87] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. *J Am Chem Soc*, (106), 1984.

[88] T. Yanai. *Chemical Physics Letters*, (393), 2004.

[89] Y. Zhao and D.G. Truhlar. *J Chem Theory Comput*, (2), 2006.

[90] Y. Zhao and D.G. Truhlar. *J Chem Phys*, (125), 2006.

[91] Y. Zhao and D.G. Truhlar. *J Phys Chem A*, (110), 2006.

[92] Y. Zhao and D.G. Truhlar. *Accounts Chem Res*, (41), 2008.

[93] Y. Zhao and D.G. Truhlar. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, (120), 2008.

[94] Y. Zhao, N.E. Schultz, and D.G Truhlar. *J Chem Phys*, (123), 2005.

[95] Y. Zhao, N. Schultz, and D.G. Truhlar. *J Chem Theory Comput*, (2), 2006.

[96] Y. Zhao, B. Lynch, and D.G Truhlar. *J. Phys. Chem. A*, (108), 2008.

[97] Z. Zsoldos, D. Reid, A. Simon, B. S. Sadjad, and A. P. Johnson. *Curr Protein Pept Sci*, (7), 2006.

[98] Z. Zsoldos, D. Reid, A. Simon, S. B. Sadjad, and A. P. Johnson. *J Mol Graph Model*, (26), 2007.